

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 09:25:29 ; Search time 27321.5 Seconds

(without alignments)
687.046 Million cell updates/sec

Title: US-09-303-518D-569

Perfect score: 897

Sequence: 1 atgttcgttaccatcag.....accgcacaatagcgtaa 897

Scoring table:

IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	897	100.0	897	6	A96536
C 2	897	100.0	349061	1	NMA222491
C 3	889.2	99.1	894	6	AX081572
C 4	889	99.1	897	6	A96534
C 5	889	99.1	15471	1	AE002530
C 6	889	99.1	349980	6	AX044034
C 7	799.4	89.1	894	6	A96538
C 8	360.6	10.3	369	6	A96532
C 9	92.2	215050	9	6	AL646057
C 10	57.6	6.4	125020	9	AF429315
C 11	55.4	6.2	13818	9	AF429315
C 12	45	5.1	2081	1	PFY14568
C 13	44.6	5.0	12832	1	AE003967
C 14	42.2	4.7	10732	6	E32986
C 15	40.8	4.5	167721	2	AC032019
C 16	40.8	4.5	168414	2	AC023146
C 17	40.8	4.5	270409	2	AC092715
C 18	40.2	4.5	125020	9	AF429315
C 19	38.6	4.3	12680	1	AE005970
C 20	38.6	4.3	166919	2	AL591675
C 21	38.6	4.3	187334	2	AL626766
C 22	38.2	4.3	10991	1	AE001874
C 23	38.2	4.3	30000	6	AX250263
C 24	38	4.2	4494	9	AF029343
C 25	38	4.2	94473	9	AL445288
C 26	37.2	4.2	138039	2	AC016786
C 27	37.2	4.1	155124	3	AB008449
C 28	37	4.1	3648	3	AC017294
C 29	37	4.1	63904	3	AC092239
C 30	37	4.1	166441	3	AC004758
C 31	37	4.1	174311	3	RME603647
C 32	37	4.1	189333	3	AE003611
C 33	37	4.1	259973	3	AP001517
C 34	36.8	4.1	294250	1	AE005014
C 35	36.6	4.1	11050	1	AE004441
C 36	36.6	4.1	12213	1	AE002369
C 37	36.4	4.1	10194	1	NMA522491
C 38	36.4	4.1	32961	1	AE001773
C 39	36.4	4.0	12514	6	AX043922
C 40	36.2	4.0	176106	2	AC018567
C 41	36.2	4.0	185139	2	AC018931
C 42	36.2	4.0	185823	2	AC021934
C 43	36.2	4.0	194763	2	AC026740
C 44	36.2	4.0	230575	2	AP001590
C 45	36.2	4.0	1312	9	HSLGUD11
C 46	36	4.0	2950	9	HSLGUD11
C 47	36	4.0	2950	9	HSLGUD11
C 48	36	4.0	2950	9	HSLGUD11
C 49	36	4.0	2950	9	HSLGUD11
C 50	36	4.0	2950	9	HSLGUD11
C 51	36	4.0	2950	9	HSLGUD11
C 52	36	4.0	2950	9	HSLGUD11
C 53	36	4.0	2950	9	HSLGUD11
C 54	36	4.0	2950	9	HSLGUD11
C 55	36	4.0	2950	9	HSLGUD11
C 56	36	4.0	2950	9	HSLGUD11
C 57	36	4.0	2950	9	HSLGUD11
C 58	36	4.0	2950	9	HSLGUD11
C 59	36	4.0	2950	9	HSLGUD11
C 60	36	4.0	2950	9	HSLGUD11
C 61	36	4.0	2950	9	HSLGUD11
C 62	36	4.0	2950	9	HSLGUD11
C 63	36	4.0	2950	9	HSLGUD11
C 64	36	4.0	2950	9	HSLGUD11
C 65	36	4.0	2950	9	HSLGUD11
C 66	36	4.0	2950	9	HSLGUD11
C 67	36	4.0	2950	9	HSLGUD11
C 68	36	4.0	2950	9	HSLGUD11
C 69	36	4.0	2950	9	HSLGUD11
C 70	36	4.0	2950	9	HSLGUD11
C 71	36	4.0	2950	9	HSLGUD11
C 72	36	4.0	2950	9	HSLGUD11
C 73	36	4.0	2950	9	HSLGUD11

A96536 Sequence 56
AL162753 Neisseria
AX081572 Sequence
A96534 Sequence 56
AE002530 Neisseria
AX044034 Sequence
A96538 Sequence 57
A96532 Sequence 56
AL646057 Ralstonia
AF429315 Homo sapi
AE004747 Pseudomon
Y14568 Pseudomonas
AE003967 Xylella f
E32986 Gene encodi
AC032019 Homo sapi
AC023146 Homo sapi
AC092715 Homo sapi
AF429315 Homo sapi
AE005970 Caulobact
AL591675 Mus muscu
AL626766 Mus muscu
AE001874 Deinococc
AX250263 Sequence
AF029343 Homo sapi
AL445288 Human DNA
AC016786 Homo sapi
AC08763 Oryza sat
AB008449 Bombyx mo
AC017294 Drosophill
AC092239 Drosophill
AC004758 Drosophill
AL603647 Rhizobium
AE003611 Drosophill
AP001517 Bacillus
AE005014 Halobacte
AE004441 Pseudomon
AE002369 Neisseria
AL162753 Neisseria
AX043922 Sequence
AE001773 Thermotog
AC018567 Homo sapi
AC018931 Homo sapi
AC021934 Homo sapi
AC026740 Homo sapi
AC026740 Homo sapi
AP001590 Homo sapi
X66300 H. sapiens G
X07769 Human mRNA
M37154 Human gluta
X07674 Human mRNA
J03248 Human liver
M20867 Human gluta
I66494 Sequence 14
AL136982 Human DNA
U12178 Human herpe
E00397 DNA coding
E00403 Herpes simp
I04285 Sequence 9
I07886 Sequence 5
X01456 Herpes simp
X01996 Herpes simp
M10053 Herpes simp
D80006 Human RNA
AE004874 Pseudomon
AL163003 Streptomy
AX110939 Sequence
X77366 H. sapiens H
AE005248 Escherich
I02373 Escherichia
AE000173 Escherichia
D90709 Escherichia
AE006946 Mycobacte
Z84774 Mycobacteri

[illegible]

QY 121 cggctcgagacatctgcygttcttacctttaagaagagacccgcgcgcgtcccaat 180
|||||
Db 121 CGGCTCGGACATCTGGCGTTTACCTTTAAAGAAAGACCGCGCATCTCGCCAAAT 180
QY 181 atgcgtcagcgagcagatgaatccgaccccaaaacggtcaaacgctttttgcygaaag 240
|||||
Db 181 ATGCGTCAGGCGAGGATGATATCCGACCCCAAAACAGTCMAAGCCGTTTGGCGGAACG 240
QY 241 gcaaaagcgggtttggaacttgcgcgcgcgtttttcaaaaaacgggaagacatagaaca 300
|||||
Db 241 GCAAAAGCGGTTTGGAACTTGGCCCGCGTTTTCAGAAAAACGGGAAGACATGAAAAA 300
QY 301 atgttcaaacgagctacacgagcttgggaacatgucagcagagcttggacaacaacgaaggg 360
|||||
Db 301 ATGTTCAAGCGGTACAGCGCTGGGAACATGTGACGAGCGTTTGGCAAAACAGGAAGG 360
QY 361 ctgcattcatcacgcgcgcacatcgagcagctacgatttggcgagagctacatcaagcag 420
|||||
Db 361 CTGCTATTTCATCACGCCGACATCGCGAGCTACGATTGGCGGAGCGCTACATCAGCCAG 420
QY 421 caagcttcggtcccgctgacgcgcacatgacaaacgcgcgaataacgaagtagaagaa 480
|||||
Db 421 CAGCTTCGGTTCCCGCTGACCGCATGTACAAACCGCGAAATCAAGCGATGACAAA 480
QY 481 atcatgcagcgagcgaggttctcgcggaagaaacacgcgcctacacgaatacaaggg 540
|||||
Db 481 ATCATGACGAGGCGGAGGCTTCGCGGCAAAAGCAAAACCGCGCTACAGCATACAAAGG 540
QY 541 gtcaaaacaatcatcaaaagcctgcttgcggcggaagaaacacatctctcgcgcgacac 600
|||||
Db 541 GTCAAAACAATCATCAAAAGCCCTGCTGCGGCAAGCAACACATCTCTGCGGAAACCTGCC 600
QY 601 gtccctccctcacaagaagcgagggaagcgatggttggatttcttcttcgcaaacctgac 660
|||||
Db 601 GTCCCTCCCTCAAGAGCGGAGGAGGCGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 tataccatgaagcttgcggcggaataatgacacagctcaaaagcggtgaacacccgttttcc 720
|||||
Db 661 TATACCATGACGCTGGCGGCAAAATTTGGCAACGCTCAAAAGCGGAAACCGCTGTTTTC 720
QY 721 tgcgtcgaaagcgtctgctgagcgaaggttgcattgtacacacgcgcgcgttccaaagg 780
|||||
Db 721 TGCTGCAAGCGCTGCTGGCGGCAAGGTTTGTGATTTGCAATCCGCCCGCTCCAAAGG 780
QY 781 gaattgaagcgagcaaaagcccatgatacgatcgcgcggtgttcaacgcgaatgcgaatatgg 840
|||||
Db 781 GAATTGAACGGCGACAAAGCCCATGATGCCGCGTGTTCAAACCGCAATGCCGAATATTGG 840
QY 841 atacgcggttttcgacgacgatactgttattgtacaaacgcgtacaaaatggcgg 894
|||||
Db 841 ATACGCGGTTTTCGAGCGAGTATCTGTTATGTATGACAAACCGCTACAAATATGCCG 894

RESULT 4
A96534
LOCUS A96534 897 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 567 from Patent WO9924578.
ACCESSION A96534
VERSION A96534.1 GI:6780159
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 897)
AUTHORS Pizsa, M., Scarlato, V., Rappuoli, R., Grandi, G. and Masiagnani, V.
TITLE Neisserial antigens
JOURNAL Patent: WO 9924578-A 567 20-MAY-1999;
PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES
Location/Qualifiers
1..897
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 223 a 266 c 227 g 181 t
ORIGIN

Query Match 99.1%; Score 889; DB 6; Length 897;
Best Local Similarity 99.4%; Pred. No. 7.7e-221;
Matches 892; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 atgttgcattacaatcaagctgttccccccttgcgaacgcgcacatgcacatcgttg 60
|||||
Db 1 ATGTTTGTATACAAATGACGCTGTTTCCCTTTGGGAACGCCCATGCATCTCTGTTG 60
QY 61 accgccttgctcaaatgacctccctcgcgtgcgcttctctgtctgcaacgcgtggaaac 120
|||||
Db 61 ACCGCTCTGCTCAAAATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
QY 121 cggctcgagacatctgcygttcttacctttaagaagagacccgcgcgcgtcccaat 180
|||||
Db 121 CGGCTCGGACATCTGGCGTTTACCTTTAAAGAAAGACCGCGCATCTCGCCAAAT 180
QY 181 atgcgtcagcgagcagatgaatccgaccccaaaacggtcaaacgctttttgcygaaag 240
|||||
Db 181 ATGCGGACGCGGCTTGAACCCGACCCCAAAACGCTCAAGCCGTTTGGGGAAGG 240
QY 241 gcaaaagcgggtttggaacttgcgcgcgcgtttttcaaaaaacgggaagacatagaaca 300
|||||
Db 241 GCAAAAGCGGTTTGGAACTTGGCCCGCGTTTTCAGAAAAACGGGAAGACATGAAAAA 300
QY 301 atgttcaaacgagctacacgagcttgggaacatgucagcagagcttggacaacaacgaaggg 360
|||||
Db 301 ATGTTCAAGCGGTACAGCGCTGGGAACATGTGACGAGCGTTTGGCAAAACAGGAAGG 360
QY 361 ctgcattcatcacgcgcgcacatcgagcagctacgatttggcgagagctacatcaagcag 420
|||||
Db 361 CTGCTATTTCATCACGCCGACATCGCGAGCTACGATTGGCGGAGCGCTACATCAGCCAG 420
QY 421 caagcttcggtcccgctgacgcgcacatgacaaacgcgcgaataacgaagtagaagaa 480
|||||
Db 421 CAGCTTCGGTTCCCGCTGACCGCATGTACAAACCGCGAAATCAAGCGATGACAAA 480
QY 481 atcatgcagcgagcgaggttctcgcggaagaaacacgcgcctacacgaatacaaggg 540
|||||
Db 481 ATCATGACGAGGCGGAGGCTTCGCGGCAAAAGCAAAACCGCGCTACAGCATACAAAGG 540
QY 541 gtcaaaacaatcatcaaaagcctgcttgcggcggaagaaacacatctctcgcgcgacac 600
|||||
Db 541 GTCAAAACAATCATCAAAAGCCCTGCTGCGGCAAGCAACACATCTCTGCGGAAACCTGCC 600
QY 601 gtccctccctcacaagaagcgagggaagcgatggttggatttcttcttcgcaaacctgac 660
|||||
Db 601 GTCCCTCCCTCAAGAGCGGAGGAGGCGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 tataccatgaagcttgcggcggaataatgacacagctcaaaagcggtgaacacccgttttcc 720
|||||
Db 661 TATACCATGACGCTGGCGGCAAAATTTGGCAACGCTCAAAAGCGGAAACCGCTGTTTTC 720
QY 721 tgcgtcgaaagcgtctgagcgaaggttgcattgtacacacgcgcgcgttccaaagg 780
|||||
Db 721 TGCTGCAAGCGCTGCTGGCGGCAAGGTTTGTGATTTGCAATCCGCCCGCTCCAAAGG 780
QY 781 gaattgaagcgagcaaaagcccatgatacgatcgcgcggtgttcaacgcgaatgcgaatatgg 840
|||||
Db 781 GAATTGAACGGCGACAAAGCCCATGATGCCGCGTGTTCAAACCGCAATGCCGAATATTGG 840
QY 841 atacgcggttttcgacgacgatactgttattgtacaaacgcgtacaaaatggcgg 897
|||||
Db 841 ATACGCGGTTTTCGAGCGAGTATCTGTTATGTATGACAAACCGCTACAAATATGCCG 897

RESULT 5
AE002530 15471 bp DNA linear BCT 25-MAY-2000
LOCUS AE002530
DEFINITION Neisseria meningitidis serogroup B strain MCS8 section 172 of 206

of the complete genome.
 AE002530 AE002098
 VERSION AE002530.1 GI:7227054
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis MC58.
 Neisseria meningitidis MC58.
 Bacteria; Proteobacteria; beta subdivision: Neisseriaceae;
 Neisseria.
 1 (bases 1 to 15471)
 Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Pedersen, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Ciftone, H., Clark, E.B., Cotton, M.D., Uterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V.,
 Pizsa, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 Complete genome sequence of *Neisseria meningitidis* serogroup B
 strain MC58
 Science 287 (5459), 1809-1815 (2000)
 2 (bases 1 to 15471)
 Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Pedersen, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Ciftone, H., Clark, E.B., Cotton, M.D., Uterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V.,
 Pizsa, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 Direct Submission
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 15471
 /organism="Neisseria meningitidis MC58"
 /strain="MC58"
 /db_xref="taxon:122586"
 /note="serogroup: B"
 78. 974
 /gene="NMB1801"
 78. 974
 /note="similar to SP:P24187 GB:X61000 PID:48957 GB:U00096
 PID:1651518 percent identity: 46.44; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="HtrB/msbA family protein"
 /protein_id="AAF42138.1"
 /db_xref="GI:7227055"
 /translation="MERDOFLPPPLRPMHLLALLKCLSLDPLSGLHTLGNRIGH
 LAEYLLEKEDRAIVANMRQAGLNPDPKTVKAVFAFTAGGDELAPFRKREDITM
 KAVGHEWQOALDKHEGLFTTPHIGSLDGRYISQOLPPLTAMKPKPTAIDK
 IMAGVRGKGTADPISIGVKQIITKALSGEATIVLDPHVSPOEGEGWAVDFGK
 PAYTTLAAKLAVHVGKTLFFCCERLPGCGFDLHPRVQGLNGDKRAHAAVFNRN
 AEWIRRPPTQYLWYRNRYKMP"
 complement(1018. .2082)
 /gene="NMB1802"
 /complement(1018. .2082)
 /note="similar to SP:P36175 PID:561690 percent identity:
 72.50; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="O-sialosylprotein endopeptidase"
 /protein_id="AAF42139.1"
 /db_xref="GI:7227056"
 /translation="MLVIGTSSCDGTGVALYDTERGLRAHCLHTOMAHAYGGVVP
 ELASRHIRLVPVLTGECIAQACASIGDIDNVAFFQGPGLGALLAGSSYANALAL

DKPVIPIVHHLEGLHLLAEKPDPEFVALLVSGGHTQIMAVRIGDYGALLGESVDD
 AAGEFDTKATLGLLYPGAKLSLAEAGREFAEVPFRPMIHSDDLQMSFGTKTAY
 LTAVEKVRARENGADDIPROTNRDLCIARQADVVVLAQVKKALLOTGTFRVYAGVY
 GANRKLRETFGNMTYQIPTPKCKPKHPSEKSVFPPPTAYCTDNGAMIAFGAMHLGK
 GREVGAFNVRPWPVLSLSEIR"
 complement(2199. .3386)
 /gene="NMB1803"
 complement(2199. .3386)
 /note="similar to GP:3790604 percent identity: 62.53;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="cytochrome c-type biogenesis protein, putative"
 /protein_id="AAF42140.1"
 /db_xref="GI:7227057"
 /translation="MTEHKTIPEHEILLQKSLISNLMMDVFAVFAVFAVYQTR
 SGKMDIYETVMLASAGTAVFLGFFRPMRFPVLSLTAAYANGYSGTKSAEIR
 LLYRFLSSQSAIMOCAYFALFAYISGAVLASVKNPNTLLMGKTVFMAVSAGV
 FGLLVRHMEVSLRPDAGHIPVSNLYEFLILEVITLALMDLYGKFAIQGLGFEV
 GMAVYVGVFLWYSVSREAHITQPLIPALQSMKTHVPANFYGAGFICSMGLIAE
 IYSLRREGSKGLMLPPSALIDEVMYKAIVGFLFTTATIGALMAADANGRYVSWD
 PKETAFPIWLNKAVAWLHLRLVAGNRGKVLAWMALLGLFVFAFIVGNMFLSGHSY
 GLT"
 complement(3379. .5394)
 /gene="NMB1804"
 complement(3379. .5394)
 /note="similar to GP:3790604 percent identity: 58.44;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="cytochrome c-type biogenesis protein, putative"
 /protein_id="AAF42141.1"
 /db_xref="GI:7227058"
 /translation="MGSRSRSPPLSRPFAFSSMRFAVALSLGLIASVIGTVLQO
 NOPQTYVYKFGSFMNOIFGFLGYVAVSAMVYVEMGFOCKTIRREGSVLIAAK
 MSKREKVRKSLAAMRHSLLDYKTAIPYAKRYLEQFOCKTIRREGSVLIAAK
 GSNKNGYIFAHVALVTCIGGLIDNSLLKGLMGLGRVYPNQOAYAKDFPESTIG
 ASNLISRGVNIISGOSADVFLNANGILVODLPEVYLLKLFHDFITGTGPRDFAS
 DIENTVRKATGEKLEERTIRVNHPLTLHGITTYASPADGSDLPFKAMNLGDSAREVY
 LKATSIHOPLEIGHKRYLEFDOFTSMNVEDSEGAERKSLKSTLNDVRYALVTEGK
 KYNIGPSIYIRIRDAAGOAVEKKNMPLVLODQDYFWLTGTRSGLOOQYRMLTIPD
 KQKAPFPALREFLIDEGSRKRLVADATKAAALADETIRRYGLAEMQDLEARRPFLHS
 LDEITISNTPKEODKMGCGYFEMLYGVNMAALDETRIRRYGLAEMQDLEARRPFLHS
 MDAYTGLTEYPAPMLLQDGEFSEVRSKSGLOMTRSPALLYVIGSVLILVGYLMTYVR
 EKRAWYLFSDGKIRFAMSSARSRDLQKPRHVESLQRLGDLNHD"
 complement(5608. .6231)
 /gene="NMB1805"
 complement(5608. .6231)
 /note="similar to GB:U05988 PID:516878 SP:Q52369 percent
 identity: 59.61; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="cytochrome c4"
 /protein_id="AAF42142.1"
 /db_xref="GI:7227059"
 /translation="MKRLLTLLAVLAAGVASAPKADYEGKQVAAATVCAACAAACGN
 SGIMYPRLLAQAHTAYLYHOTIGIRDKRRHGSAAVKKPVVNMISDDILLNVAFTAK
 QQPKSGANPRENELGAKITRYRGLISDKVPACMSCHPGSGAGMPGGSSIDVAPPLG
 GQHOAYIVEQNMVYKSGQRKNTIMEDIANRMSDDLKAVANFIOGLR"
 6431. 7066
 /gene="NMB1806"
 6431. 7066
 /note="conserved hypothetical protein; identified by
 glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF42143.1"

/db_xref="GI:7227060"
/translation="MPMNLFONAKFTTINHLKDLPTPLEIAFVGSNKSNAINT
LTHNVLAVYSKTPGTROHINFEELONGNVMVDPGVYQVQVEAVAHVNNILGTYL
QORROLIGVILINDARHPLKEDIRMDPFHTTGRPHILSLKADLSKNEQITLSQ
VKLLKISYSDRONISVOLSESLKKQGIIDEANRTVGSWLDADADAASSPEEN"
complement(7187, .9583)
/gene="NM1807"
/complement(7187, .9583)
/note="Similar to PID:1916074 percent identity: 100.00;
identified by sequence similarity; putative"
/transl_table=11
/product="penicillin-binding protein 1"
/protein_id="AA042144.1"
/db_xref="GI:7227061"

/translation="MIKILITGCGVFCVGVGLVAIIIVTTPKLPDSLSLOH
OAKMPLTISADGEVIMGMEQRRPEFKIDPEVLEKNVIAEDKRFYRHMGVWG
VARAANGNVSSVSGASTTQVAKNFWLSEKTEFRNFVLLAYKIEQSLKDK
ILIELFNQYILGQRAFGASAAQIYFNKNVRLTLAFAAMACLPAPASVYNPVNE
RAKLROKYLINMLIEKMITVQDQALNELHYEFERKIIDQSALYVEMVROELEY
KVEDAVYTOGFKVTYVRADHQRATELALRNDRGSRVGAENYIDLSKDEVE
ETVSYLSGLYTDKMPAVYLDVTKKNVYIOLPGRRVITDRALGFAARVYNEK
MGEDRIIRGAVIRVKNNGRMNAVOPERLQALVSDATGAVRALVGGIDHSKTFN
RAYOAMROESTKPPYTSKALSKGMTASTVVDADISLPGKGNPSVMTPKNSDERY
SGITTLKQALTSKNNVSIKILMSIGVAAQYIRFRGRSSSLPALSMAIGTGETT
PLKVAEAYSVFANGYRVSVSHVVDKIDYDRGRRLAOMPLVAGONAPQALDPNNATIM
YKIMOVYGVGTARGAALGRDIDAGKTGTNNKRNKAWFVGVDPVTAVYIGDFPK
SMGRVYGVGTIAVPMVYVAFALGCKQCKMMPGVYSNENYEMKRMVYDRLT
LDNSGIAPQPSRRAKEDDGAAGGRQADDEVRQDMQETPVLPSNTGSKQOOLDSLF
"

gene
9735, .10850
/gene="NM1808"
9735, .10850
/gene="NM1808"

/note="similar to PID:1914831 percent identity: 98.11;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="p11m protein"
/protein_id="AA042145.1"
/db_xref="GI:7227062"

/translation="MRLPKSLKNPKTKDAKLPRKSSGLNNRAIISIDIDSHIKVOL
SGRSLNQLDEKVIAPKIPKNTIQKKNVNDQVLYLQAAAKLIGTSKNTIASVPO
NLATIEQLTYTDKDELIDQGFVESISEVSSISLEENYDVQVLSQSAAGVAVLA
SKRDELEPLIDAFNAAGMKLSALVDIVFQYNAVALMHNHFAPELAEKAVAFGYAA
QYVALVIOGKILYKQETSVESEOLNLIORTYOVTEKEAEETINSPOKPSYQESVA
NYENQITQEIORVLOFYTTOTADMDMDIKIILITGEAARQEGIAQVYASQTNADVQ
CVHPARVADNPKTDKQPELDAPLTLTRAFGLAVNGL"
10853, .11452
/gene="NM1809"
/note="This region contains an authentic frame shift and

Query Match 99.1%; Score 889; DB 1; Length 15471;
Best local Similarity 99.4%; Pared No. 8 7e-221;
Matches 892; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 atgttcgtttacaattcagcgtgtttccccccttgcagaccgcatcactcgttg 60
DB 78 ATGTTTCGTTTACAAATGAGGCGTGTTCCTCCCTTCGAAACGCCATGACATCCGTG 137
QY 61 accgcctctcaaatgctctccctgcgcgtcttctctgtctgcagcgttggaagc 120
DB 138 ACCGCGCTGCTCAAAATGCTCTCCCTGCGCGCTTCTGTCACACGCTGGGAAAC 197
QY 121 cggctcgaacatcgcgtttacaccttaaggaagaccgcgcgcgcgcgcgcgcgcgc 180
DB 198 CGGCTCGGACATCTGGCGTTTACCTTTAAAGAAAGACCGCGCGCATCTGCGCAT 257
QY 181 atgcgtcgaagc 240
DB 258 ATGCGGACGAGCGGCTTTGAACCCGCAAAACGCTCAAGCGCTTTTTCGGAACG 317

QY 241 gcaaaagcggttttggaacttgcctcccggtttttcagaanaaccggaagacatagaaca 300
DB 318 GCAAAAGCGGTTTGGAACTTGGCCCCCGCTTTTTCAGAAAACCGGAAGACATAGAACA 377
QY 301 atctcaaacggttacacgccttggaacatgttgcagcagcgtttgacaacaacggaag 360
DB 378 ATGTTCAAAACGGTACACGGCTGGGAACATGTGACACAGGCTTTGGACAAACGGAAGG 437
QY 361 ctgctattcacaagcgcgcacalcgcagctacagatlttgaggcagcgtacacacagcag 420
DB 438 CTGCTATTCAATCACCGCGCATCGCAGCTAGATTTGGCGGAGCGTACATCAGCCAG 497
QY 421 caacttcggttcccgctgcagcgcacgtacaaacgcgcggaataaacaagcagtagaaca 480
DB 498 CACCTTCGTTTCCCGCTGACCGGCATGTCAAAACCCCAAAATCAAAACGATAGACAA 557
QY 481 ataatcgaagc 540
DB 558 ATCATGCAAGCGGCGGAGGCTGCGGCAAAAGAAAACCGCGCTACACATACAGAGG 617
QY 541 gtaacaacaatcacaagccttgcgttcgggaggaagcaacacatcgtcctgcgcagcac 600
DB 618 GTCAAAACAATCATCAAAACCCCTGCTGCGGGAAGCAACCATGCTGCTGCGCGACAC 677
QY 601 gtccctccctcgaagaagc 660
DB 678 GTCCCTCCCTCAAGAACGCGGAGCGGATGGGTGATTTCTTGGCAAAACCTGCTC 737
QY 661 tataccaagcgttcgc 720
DB 738 TATACCATGACGCTGCGGCAAAATTTGGCACAGTCGAAGGCGTGAATAAACCTGTTTTC 797
QY 721 tcttggaagc 780
DB 798 TCTTGCGAAGCGCTGCTGCGGACAGGCTTGCATTTGCAATCCGCCGCTCAAGAGG 857
QY 781 gaattgaacgc 840
DB 858 GAATTGAAGCGGACCAAGCCCATGATGCGCGCTTCAACCGCAATGCCGAATATTGG 917
QY 841 ataagcgttttcgc 897
DB 918 ATACGCGCTTTCGACGACGATCTGTTAGTACAAACCGGTACAAATGCCGTAA 974

RESULT 6
AX044034
LOCUS AX044034
DEFINITION Sequence 113 from Patent WO0066791.
ACCESSION AX044034
VERSION AX044034.1 GI:11342918
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 349980)
Piazza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C.,
Maignan, V., Galeotti, C., Mora, M., Ratti, G., Scariselli, M.,
Scarlatto, V., Rappunli, R., Frazer, C.M. and Grandi, G.
Neisseria genomic sequences and methods of their use
Patent: WO 0066791-A 113 09-NOV-2000;
CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
location/qualifiers
1..349980
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
/note="sequence too long, cut in 8 pieces. -seq 1: 1 to
349980 349980 bases-seq 108: 300001 to 649980 349980
bases-seq 109: 600001 to 949980 349980 bases-seq 110:
900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to

2272325 172325 bases"
 BASE COUNT 86473 a 95646 c 85908 g 81953 t
 ORIGIN

Query Match 99.1%; Score 889; DB 6; Length 349980;
 Best Local Similarity 99.4%; Pred. No. 1e-220;
 Matches 892; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 atgttcgtttacattcaatgagctgttcccttgcgaacgcagcatcattccgtgtg 60
 |||||||
 Db 89000 ATGTTTCGTTTACAAATTAAGCTGTTTCCCTTCCCAACGCGCATCTCTGTTG 89059
 QY 61 accgcctcgtctcaaatgctctcctcgtcgcgtcttcctgtcgtacacgcgtggaaac 120
 |||||||
 Db 89060 ACCGCCCTGCTCAAAAGCCTCTCCCTGCTGCGCTTCTCTGCTGCACACGCTGGGAAC 89119
 QY 121 cggctcggacatctggtcgttttaacctttaagaaagccggcgagcatcgtccaat 180
 |||||||
 Db 89120 CGGCTCGACATCTGGCGTTTACCTTTTAAAGAAAGACCGCGCGCATCTGCTCCCAT 89179
 QY 181 atgcgtcagcagcagcatgcatccgaccccaaacggtcaaaacgcttttgcggaacg 240
 |||||||
 Db 89180 ATGCGGACAGCGGCTTTGAACCCCGACCCCAAAAGCGTCAAAAGCGTTTTCGGAACG 89239
 QY 241 gcaaaagcgttttggaacttgcgcgcgttttccagaagaacccggaagacataagaac 300
 |||||||
 Db 89240 GCAAAAGCGGCTTTGGAACCTTCCCGCGTTTTCAGAAACCGGAACACATAGAAACA 89299
 QY 301 atgttcaagcgtatcacagcgtcggagacatgtcgcagcagcttttgacaacaacgaaagg 360
 |||||||
 Db 89300 ATGTTTCAAAAGCGGTACACGCGTGGGAACATGTGACAGCAGCTTTGGACAAACACGAAGG 89359
 QY 361 ctgtctatcatcaagcgcagacatcggcagctcagatttgcggaagcgtacacacacg 420
 |||||||
 Db 89360 CTGCTATTATCATCAGCGCGACATCGGCACATGATTTGGGGAGGAGCTACATCAGCCAG 89419
 QY 421 cagcttcggttcccgctgcagcagcatgtatcaaacgcgcaaatcaaacgataagaac 480
 |||||||
 Db 89420 CAGCTTCCGTTCCCGCTGACCGCGCATGTACAAACCCCGAAATCAACGATAGCAAA 89479
 QY 481 atcatgcaagcggcaggggttcgcggaaggaagaaacgcgctcaccagatcaagaagg 540
 |||||||
 Db 89480 ATCATGAGGCGGCGGAGGGTTCGCGCAAAAGCAAAACCGCGCTACGACATCAAGGG 89539
 QY 541 gtaacaacaatcatcaagccctgcgttcggggaagcaaacatgctctccggagacac 600
 |||||||
 Db 89540 GTCAAAACAAATCATCAAAAGCCTGCTGCGGAGCAACCATGCTGCGGACACAC 89599
 QY 601 gtcccccctcaagaagcggggaagcgatattggttgatttcttcggaacactgc 660
 |||||||
 Db 89600 GTCCCTCCCTCTCAAGAAGCGGGGAAGGCGTATGGGTGATTTCTTGGCAAAACCTGCG 89659
 QY 661 tatacatalgaagctgcgcaaaatttgcacacgtcaaaagcggtgaacacccgttttc 720
 |||||||
 Db 89660 TATACCATGACGCGTGGGCAAAATTTGGCACAGCTCAAAAGCGTGAACCCCTGTTTTC 89719
 QY 721 tgcctggaagcgtgcctgcggaagcggttcgatttcacatccgccccgtccaagg 780
 |||||||
 Db 89720 TGGTGGCAAGCGCTGCTGGGCAAGGTTTGGATTTCACATCCGCCCGCTCCAAAGG 89779
 QY 781 gaattgaagcgcaaaagcccatgtatgcgcgcgtgttcaacgcgaatgcgaatatgg 840
 |||||||
 Db 89780 GAATTGAAGCGGCAAAAGCCCATGATGCGCGTGTTCAAACGCAATGCGGAATTTGG 89839
 QY 841 atacgcgttttccgagcagatctcgttattatgataacgcgtacacaaatgcggtta 897
 |||||||
 Db 89840 ATACGCGGTTTCCGACGAGATCTGTTATGATACAAACGCTTACAAATGCGGTAA 89896

RESULT 7
 A96538
 LOCUS A96538 894 bp DNA linear PAT 07-SEP-2000

DEFINITION Sequence 571 from Patent WO924578.

ACCESSION A96538
 VERSION A96538.1 GI:6780161

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Score 799.4; DB 6; Length 894;

Pred. No. 1.7e-197;

Mismatches 51; Indels 3; Gaps 1;

QY 1 atgttcgtttacattcaatgagctgttcccttgcgaacgcagcatcattccgtgtg 60

Db 1 ATGTTTCGTTTACAAATTAAGCTGTTTCCCTTCCCAACGCGCATCTCTGTTG 60

QY 61 accgcctcgtctcaaatgctctcctcgtcgcgtcttcctgtcgtacacgcgtggaaac 120

Db 61 ACCGCCCTGCTCAAAAGCCTCTCCCTGCTGCGCTTCTCTGCTGCACACGCTGGGAAC 120

QY 121 cggctcggacatctggtcgttttaacctttaagaaagccggcgagcatcgtccaat 180

Db 121 CGGCTCGACATCTGGCGTTTACCTTTTAAAGAAACCGCGCGCATCTGCTCCCAT 180

QY 181 atgcgtcagcagcagcatgcatccgaccccaaacggtcaaaacgcttttgcggaacg 240

Db 181 ATGCGGACAGCGGCTTTGGAACCTTCCCGCGTTTTCAGAAACCGGAACACATAGAAACA 240

QY 241 gcaaaagcgttttggaacttgcgcgcgttttccagaagaacccggaagacataagaac 300

Db 241 GCAAAAGCGGCTTTGGAACCTTCCCGCGTTTTCAGAAACCGGAACACATAGAAACA 300

QY 301 atgttcaagcgtatcacagcgtcggagacatgtcgcagcagcttttgacaacaacgaaagg 360

Db 301 ATGTTTCAAAAGCGGTACACGCGTGGGAACATGTGACAGCAGCTTTGGACAAACGGAAGG 360

QY 361 ctgtctatcatcaagcgcagacatcggcagctcagatttgcggaagcgtacacacacg 420

Db 361 CTGCTATTATCATCAGCGCGACATCGGCACATGATTTGGGGAGGAGCTACATCAGCCAG 420

QY 421 cagcttcggttcccgctgcagcagcatgtatcaaacgcgcaaatcaaacgataagaac 480

Db 421 CAGCTTCCGTTCCCGCTGACCGCGCATGTACAAAGCGCGGAATCAAAAGCGATAGACAAA 480

QY 481 atcatgcaagcggcaggggttcgcggaaggaagaaacgcgctcaccagatcaagaagg 540

Db 481 ATCATGAGGCGGCGGAGGGTTCGCGCAAAAGCAAAACCGCGCTACGATCAAGGG 540

QY 541 gtaacaacaatcatcaagccctgcgttcggggaagcaaacatgctctccggagacac 600

Db 541 GTCAAAACAAATCATCAAAAGCCTGCTGCGGAGCAACCATGCTGCGGACACAC 600

QY 601 gtcccccctcaagaagcggggaagcgatattggttgatttcttcggaacactgc 660

Db 601 GTCCCTCCCTCTCAAGAAGCGGGGAGGCGTATGGGTGATTTCTTGGCAAAACCTGCG 660

QY 661 tatacatalgaagctgcgcaaaatttgcacacgtcaaaagcggtgaacacccgttttc 720

Db 661 TATACCATGACGCGTGGGCAAAATTTGGCACAGCTCAAAAGCGTGAACCCCTGTTTTC 720

Oy 721 tgccttcgaacagcgtcgtctgagcgagcaaaagtttcgatttgcacatccgcgcccttccaaagg 780
 Db 718 tctcttggaacgccttgccgccacgcgacaaggcttctgtttgcacatccgcgcccttcgcaagg 777
 Oy 781 gaattgaacgcgagcaaaagcccatgatgcgcgcgtgttcaaccgcaatgcgaaatttg 840
 Db 778 gaattgaacgcgacaacaaaagcccatgatgcgcgcgtgttcaacgcgaattccgaattttgg 837
 Oy 841 ataagcgttttccgacgcagatctgttattgtacaacgcgttacaaacgtttaaagtcgftaa 897
 Db 838 ataccgccttttccgacgcagatctgttattgtacaacgcgttataaaaaagccgftaa 894

RESULT	8				
A96532					
LOCUS	A96532	369 bp	DNA	linear	PAT 07-SEP-2000
DEFINITION	Sequence 565 from Patent WO924578.				
ACCESSION	A96532				
VERSION	A96532.1	GI:6780158			
KEYWORDS					
SOURCE	unidentified.				

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 369)	Pizza, M., Scariato, V., Rappuoli, R., Grandi, G. and Masiugnani, V.	Neisserial antigens	Patent: WO 9924578-A 565 20-MAY-1999;

CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
Location/Qualifiers

	BASE COUNT	ORIGIN
/organism="unidentified"	84	a
/db_xref="taxon:32644"	107	c
	93	g
	84	t
		1 others

Query Match	40.2%	Score 360.6;	DB 6;	Length 369;
Best Local Similarity	98.4%;	Pred. No. 3.5e-83;		
Matches 363; Conservative	1;	Mismatches 5;	Indels 0;	Gaps 0

Qy 1 atgttctcgattacaattcagagctgtttccccccttgcgacgcgcacatgcacatcctgttg 60
|||||
Dy 1 atgttctcgattacaattcagagctgtttccccccttgcgacgcgcacatgcacatcctgttg 60
|||||
Db 1 atgttctcgattacaattcagagctgtttccccccttgcgacgcgcacatgcacatcctgttg 60
|||||

61 accgacctgtcaaatgtccctctccctgtgcgcgttctctgtcacacgctggaac 12

121 cggctcgagacatctgcgctttacacctttaagagaagacgcgcgcgcatctgtccaat 18

121 CGGCTCGACATCTGGCGTTTACCTTTAAAGGAAGACCGCGCGCATCTCGCCMAT 18

181 ATGCGGACGCGGTTTGAACCCGACCCAAACGGTCAACGCCGTTTTCGCGGAACG 24

241 gcaaaagcggttctggaactlgcccccgcgtttctcagaataacgcgaagacatagaaca 30
|||||
241 ccaaaccc

241 GCAAAAGCCGGTTTGACCTTCCCCCGCGTTTTCAGAAACCGAAGACATGAACA 30

301 ATGTTCAAGCGGTACACGGCTGGGACATGTGCAGCAGCGCTTTGGACAAACACGAMGGG 36

```

27      ccgcctatcc 369
      |||||
Db      361 CTGCTATTC 369

```

RESULT	9			
AL646057				
LOCUS	AL646057	215050 bp	DNA	linear
				ECT 07-DEC-2001

DEFINITION	Accession	Version	Keywords	Source	Organism
Ralstonia solanacearum segment 1/19	AL646057	AL646052	AL646057.1	GI:17427008	Ralstonia solanacearum.
					Ralstonia solanacearum.

REFERENCE	AUTHORS
1 (bases 1 to 215050)	Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, C.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Alaiet, M., Billault, A., Brottier, P., Camus, J. C., Catolico, L.,
 Chander, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N.,
 Gaspin, C., Layle, M., Moisan, A., Robert, C., Saurin, W., Schex, T.,
 Signier, P., Trebault, P., Whalen, M., Wincker, P., Leyry, M.,
 Weissenbach, J. and Boucher, C. A.
 Genome sequence of the plant pathogen *Ralstonia solanacearum*
 2 (Phases 1 to 215050)
 Boucher, C. A.
 Direct Submission
 Submitted (05-Dec-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston

COMMENT

FEATURES

Location/Qualifiers

<http://sequence.loulouse.inra.fr/R.solanacearum.html>.

source

rep_origin

gene

CDS

qem

CDS

```

718. .1113
/ gene="rnpA"
/ note="RSC0002; RS01824"
718. .1113
/ gene="rnpA"
/ EC_number="3.1.26.5"
/ function="macromolecule metabolism; macromolecule
degradation; degradation of rna"
/ note="Product confidence : probable
Gene name confidence : probable
predicted by Homology
predicted by Framed"
/codon_start=1

```

```

/transl_table=11
/evidence=not_experimental
/product="PROBABLE RIBONUCLEASE P PROTEIN COMPONENT"
/protein_id="CAD13530.1"
/db_xref="GI:17427010"
/translation="MTQDESSVFALRPVRSRHVELVVRANGHPQARLGIIVIGKFA
RRAERURLIRQCRLEFRLROALGGSDVLRLQKPPREDVPVAFAFKRLCREELSY
LFEIAPRLAPVPPCPAARVPDSCAAP"
1110..1424
/gene="RSC0003"
/note="RS01825"
1110..1424
/gene="RSC0003"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="CONSERVED HYDROTHERMAL PROTEIN"
/protein_id="CAD13531.1"
/db_xref="GI:17427011"
/translation="MPEVFLFLRVYKVAFSPVGNOCFLPTGCSVDARDAVLTHGPA
IGSYLAARLCRCRHPAQGGYDVPVPAPAGDAALRLSTDSASTETTPAADAAPAARPSILP
RP"
1431..3092
/gene="RSC0004"
/note="RS01826"
1431..3092
/gene="RSC0004"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : probable
gene name confidence : hypothetical
predicted by Homology
predicted by Framed"
/codon_start=11
/transl_table=11
/evidence=not_experimental
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD13532.1"
/db_xref="GI:17427012"
/translation="MDIKRTILWYIFSLSVLLTFDMNQRANGSHSQMFFPQTVTTTH
AAGGCPADVADKAAAPAAASQAAPATGAVSOTPASEKIVTVDTIRATVDPAALV
TKLELTDQDHGNGNAPFDSLSERTLARGSLIGDGPNNHTVFASGPDLTGGTGGV
EVALITLANKGAKLAKTYRKRSYVIDRPEVDNNGAPINPLTYMLADGGADV
QSFYSTFGPNAVYDTHDHRKHTFPADISKAHVAPFDPGVAVMQVYFASAPIR
ASAKREFYVDRIDTNEFYRNGAALGVAPASVASARLFAQDQBERLSEITTGIL
LVNDYQWMLTILKKPRLKLEIKHLKGLMWGMSIVALVIVKLFEPFLSTYSRSMAK
KLDVQPMPTAIREHKGDPQKQKQEMKLTLETEKVNPLGGCLPIVQIPVIALVWVL
SYVEMGACWIGDVMHDIASDPDYTLIPILAMVSFVOTRLNPPPPVPAKMMMPMP"
ASVEMFFPFPAGCIYLWVYNNCLSLAQQMSINRLGINKKAAAK"
3222..4667
/gene="thdF"
/note="RS00005; RS01827"
3222..4667
/gene="thdF"
/function="cell processes; protection responses;
detoxification"
/note="Product confidence : probable
gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE THIOPHENE AND FURAN OXIDATION PROTEIN"
/protein_id="CAD13533.1"
/db_xref="GI:17427013"

```

[illegible]

Qy	655	ccgcctataccatgcagcgtcgggcgcaaatlgtgcacagctcaaggcgtgaaacccgtc	714
Db	17596	MTGGMMRSKYYTYTGRRMMATYTCGCCWRRRSYTYRSMGMRKSSMSGMRMGSSASS	17655
Qy	715	ttttcttcgtcgcagacgcctctcgctggcggaagtttcgattgcacatccgccctgc	774
Db	17656	RCKSASMSMCSMRKMRKMSRCSKSMKMGSSRSRSASMSCKSGRGRRRSRKSSRYKR	17715
Qy	775	caaggggaaattgaacgcgcacaaagcc	801
Db	17716	GRGKRSMTKSGSKGSKCWMKRSMS	17742
RESULT 11			
AE004747/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL	</		

gene
CDS
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06629.1"
/db_xref="GI:9949364"
/translation="MSDPLAPRPRLTGLTSGGGAARAVYGVLAIAIDLLPDSAGNP
EPVIVGTSAALNAVALACGALQFROAIDRLTAWMOGRTEHYVSIASDMPGYTRQASR
LGISLITGAGRRVALLDNRPPLHELLARELDSGLHAARQOLRAAVVLAFCGESQ
AAVEYGRATIDEMVRRRIIGITRLERLHLASAAILPLPEPVRINREYFGDQAVRS
CAPISTPAKGLARVLVYISGVNPAFGSANTGLPPSPRPRLAIOGAHLINFTFID
SLERDIFOLDRMNOIGOLVAPFRRQAGGLEVEVLVSPSPDEIAARYHRLTSRV
LRLRGKPGATRASGAGVSYLLFERGYSLELIELGYDHAMGRTELEBRFLIER"
complement(2259. .3197)
/gene="PA3242"
complement(2259. .3197)
/gene="PA3242"
/codon_start=1
/transl_table=11
/product="probable lauroyl acyltransferase"
/protein_id="AAG06630.1"
/db_xref="GI:9949365"
/translation="MDPRPSSAFLLHPRYWPLFGILMLVVOLEPYVYLMLGRLG
ALMRYVSGREHLAARNLCECPKESPAEERELKFNFASSGIAFFEMAMSWMKAR
LARIALLIEGLHLEAQAQCEGYILMAHFTTEIGAILGVQHTIDGVI RHDPNV
DYVORRERERNLIDATAIEREDVYSMLKULRGRAIIVNAPDDQYGKOSLFLPLGIP
AAVATATKCAKGLRRARVLPPTQSRLDGSGYRLTIHPLEDPGSESEDLRLINOM
VERCVRQOPEQVYLAHRRFKTRPPGEAKYIPKRRK"

gene
CDS
3357. .4148
/gene="minc"
/note="PA3243"
3357. .4148
/gene="minc"
/codon_start=1
/transl_table=11
/product="cell division inhibitor Minc"
/protein_id="AAG06631.1"
/db_xref="GI:9949366"
/translation="MSQADLLDDDPYFOLKSGMLAVTILELANDLARELOLAKVA
OAPNFRPTPLVMALDKLPEGEGRIDLPALLEVCRRGLTTLAIRGREDIDIAAGAL
DLPIVPEGAREERPLDIKDSAPRKPAEPPSGEAPPEPAKAEKPAEVPSPKGV
KTPVNGMGQVIAAGSDILVLAASVSGAELLADGNIHVYGMGRALAGVGAATARIF
CQQLAAVLSTIAGNRYKAEDLRSPQNGKAVHVSLSGDVNIITRL"

gene
CDS
4210. .5025
/gene="mind"
/note="PA3244"
4210. .5025
/gene="mind"
/codon_start=1
/transl_table=11
/product="cell division inhibitor Mind"
/protein_id="AAG06632.1"
/db_xref="GI:9949367"
/translation="MAKILVVTSGKGVGTTTSAITGTLALRGFTVIVDPDVGRL
NIDILMGCERRVYDFPVNNGEATILGALKDRLKLNHLASOTRDALTKBEV
EYVMELEKRDIEYITCDSPASIEKGSAHLAMAFPAEIVVTVNPEVSSVSDSDRMGLLA
KSSQAEGKEDEYIEKHLILTRINPERYTKGEMLVDDVEELAIRLIGLVPEGOAVLK
ANQCVAVTIDEOGDAGQAYSDAVDRLLGKEIPHRFLDVQAKGFLORLPGGR"

gene
CDS
5022. .5276
/gene="mine"
/note="PA3245"
5022. .5276
/gene="mine"
/codon_start=1
/transl_table=11
/product="cell division topological specificity factor
Mine"
/protein_id="AAG06633.1"
/db_xref="GI:9949368"
/translation="MSLIDFFSRKSSNSASIAKERDIIYAHERGQRAQPDYLPQQL
KDLLEVIKRYVVIDEOIQVELENGNSIILENTTLPDR"

gene
CDS
5380. .6015
/gene="rlua"
/note="PA3246"
5380. .6015


```

gene
  /db_xref="taxon:294"
  complement(1..758)
  /gene="htrb"
  complement(<1..758)
CDS
  /gene="htrb"
  /codon_start=1
  /transl_table=11
  /protein_id="CAA74897.1"
  /db_xref="GI:2330641"
  /db_xref="SPTREMBL:O33437"
  /translation="MDKLKALVGLALFALLPRAVAVGASIGFMKLPNRSRD
  VLRINLACPEMPDARERLVGLSNDIGSLSESACAMIPARSIDVREVEGLE
  VLKELASGKGVCTSLGEMVNLNHYGSCDPIIFRPPKLADELIRKRVOL
  GNKVAATREGILSTYIKRKGGOYGFADPEPASAIFPFPAITQALISKPEPNL
  AGKAVGVFLHALRLPDGSGYKVLLEAEPMYST"
  complement(802..1359)
  /gene="tag"
  /complement(802..1359)
CDS
  /gene="tag"
  /codon_start=1
  /transl_table=11
  /protein_id="CAA74898.1"
  /db_xref="GI:2330642"
  /db_xref="SPTREMBL:O33438"
  /translation="MPCFMCSDPLVMAHYHQEWGTPLRDAQGLFELLLEGRQAGL
  SWIYLRKRRERYRVLGPDVORVAMSDAEIDELMDPGIIRNLKTKARRNAQAM
  LALDEPVFLMSFVVDKPIINFKRTEVPATTPALAMSGKKAGFTFVPTCYA
  LMAAGMVMHDHDDCDRIAEALVNGC"
  1441..2081
  /gene="g1y0"
  /gene="g1y0"
  /gene="g1y0"
  /gene="g1y0"
  /codon_start=1
  /transl_table=11
  /product="g1y0 tRNA-synthetase"
  /protein_id="CAA74899.1"
  /db_xref="GI:2330643"
  /db_xref="SPTREMBL:O33439"
  /translation="MSQPPAVTFODLILALQOYMAEGCVLQPYDMVGAGTFFHT
  ATRFLIGDETNMAAYVOPSRPDGNGENRLOHYOFQVYLKPNPEOELYLQ
  SLKHVGLDPLVHDIRVEDNENMSPILGAMGLMEVNLNHYGSCDPIIFRPPKLA
  VTELTLYGERLAMIQGVDSYDLYMADPPKGYTGIVFQHNVEQSTLEDP"
  1441..2081
BASE COUNT      406 a      643 c      632 g      400 t
ORIGIN
Query Match      5.1%; Score 46; DB 1; Length 2081;
Best Local Similarity 46.9%; Pred. No. 0.35;
Matches 225; Conservative 0; Mismatches 240; Indels 15; Gaps 2;

```

```

OY 673 ctggcggcaaatggacacgtcacaagcgtgtaaaacccgtgtttctctgctgcaagc 732
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 TTCCTACCGAATCATGCTCGCGCGCGCAAMAGCGGTGGCGTCTTCCCTCCACCGCCCTGCGC 63
OY 733 ctgcctgctgcaagcaggttgcattgacatccgcccgctcccaaggaattgacagc 792
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 CTGCGGAGCGGCTCCGCTACAAAGTCATCTCGAAGCCCGCGGAGCGATGTACGC 3

RESULT 13
AE003967/c
LOCUS
DEFINITION
Xylella fastidiosa 9a5c, section 113 of 229 of the complete genome.
AE003967 AE003849
VERSION
Xylella fastidiosa 9a5c.
Xylella fastidiosa 9a5c
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
REFERENCE
1 (bases 1 to 12832)
Simpson A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,
Alvarenga, R., Alves, L.M., Araya, J.E., Bala, G.S., Baptista, C.S.,
Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.,
Bueno, M.R., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carier, H.,
Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M.,
Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H.,
Faccinani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A.,
Fraga, J.S., Franco, M.C., Franco, M.C., Frohme, M., Furlan, L.R.,
Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A.,
Ho, P.L., Hohnsels, D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P.,
and Marino, C.L.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis
Nature 406 (6792), 151-157 (2000)
20365717
10910347
2 (bases 1 to 12832)
Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,
Alvarenga, R., Alves, L.M.C., Araya, J.E., Bala, G.S., Baptista, C.S.,
Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S.,
Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M.,
Carrier, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,
Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,
Docena, C., El-Dorri, H., Faccinani, A.P., Ferreira, A.J.S.,
Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franco, S.C., Franco, M.C.,
Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S.,
Gomes, S.L., Gruber, A., Ho, P.L., Hohnsels, J.D., Junqueira, M.L.,
Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F.,
Lambais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A.,
Lopes, R.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N.,
Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,
Martins, H.M.F., Matsukuma, A.Y., Moon, D.H., Nagai, M.A.,
Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, J.F., Nobrega, F.G.,
Nascimento, A.L.T.O., Netto, L.E.S., Nham, J.F., de Oliveira, R.C.,
Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, A.G., Pereira
Palmeiri, D.A., Paris, A., Peixoto, R.B., Pereira, G.A.G., Pereira
Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V.,
de M. Rosa, A.J., de Rosa, J.F., de Sa, R.G., de Sa, R.G., da Silva, A.M., Silva
Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva
Jr., W.A., da Silveira, J.F., Silvestri, M.L.B., Siqueira, W.J., de
Souza, A.A., de Souza, A.P., Terenzi, M.F., Trufi, D., Tsai, S.M.,
Tsubako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S.,
Vettore, A.L., Zago, M.A., Zatz, M., Zedler, J., and Setubal, J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil
Location/Qualifiers
1..12832
/organism="Xylella fastidiosa 9a5c"
FEATURES
Source

```


Db	505	CGCGGAAGGTATTTGACACATCTTCCTGGCATGGAAGCATTAACATGACTCTTGACA	446
Oy	681	aaatctggcacagctcaaaagcggtgaaaacccctgttttctgtcgggaagccgctcg	740
Db	445	GCGTATTAGCGGAGACGACACCGACACCGTCTATACCCCTGGTGGCGACCAATTAGTCC	386
Oy	741	cggacgaaggttcattgtacatc--cgcccgctccaaggggaattgaagcgcgaca	797
Db	385	TCATCTAGAGTTCCCTTGCAATATGGAACACGACCGACCGTGGCGGATCCAGATCC	326
Oy	798	agcccatgatgcgcgcgtgttcaacccgaatgcgcgaatatgatacgcgttttcgac	857
Db	325	ACTGATCCCGCGCGACCGACCTTAATGCAAGGATGCAAGTATCGCCACGCGACCCAC	266
Oy	858	gcagatcctgtttatgtacaacgcgtacaatatg	892
Db	265	TCAGTATAGTGGACGACTACAAAGCGGTAAAGCTGC	231
RESULT	14		
LOCUS	E32986	10732 bp	Linear
DEFINITION	Gene encoding cellulose synthetizer.		
ACCESSION	E32986		
VERSION	E32986.1	GI:13022340	
KEYWORDS	JP 2000060568-A/1.		
SOURCE	Vigna angularis.		
ORGANISM	Vigna angularis.		
REFERENCE	1 (bases 1 to 10732)		
AUTHORS	Koichi, M.T. K. and Sato, D. S.		
TITLE	Gene encoding cellulose synthetizer		
JOURNAL	Patent: JP 2000060568-A/1 29-FEB-2000.		
COMMENT	KOICHI MIZUNO, MITSUO GIYOUSAI SHOKUBUTSU BIO KENKYUSHO.		
	OS Vigna angularis		
	PN JP 2000060568-A/1		
	PD 29-FEB-2000		
	PE 26-AUG-1998 JP 1998239998		
	PR		
	PI KOICHI MIZUNO, TOMOHIKO KANO, SHIGERU SATO, DAISUKE SHIBATA		
	C12N15/09, A01H5/00, C07K16/40, C12N5/10, C12N9/10, C12P21/02// PC		
	(C12N15/09, C12R1:91), (C12N5/10, C12R1:91), (C12P21/02, C12R1:91),		
	C12N15/00,		
	PC C12N5/00, (C12N15/00, C12R1:91), (C12N5/00, C12R1:91) CC		
	PH Key	Location/Qualifiers	
	FT CDS	(1)..(3375).	
FEATURES	source	Location/Qualifiers	
	1..10732	Location/Qualifiers	
	/organism="Vigna angularis"		
	/db_xref="taxon:3914"		
BASE COUNT	3149 a	1212 c	2074 g
ORIGIN	2046 t	2251 others	
Query Match	4.7%;	Score 42.2;	DB 6;
Best Local Similarity	13.8%;	Pred. No. 3.7;	242;
Mismatch 75;	Conservative 224;	Mismatch 1;	Gaps 1
Oy	147	tttaagaagaagcgcgcgcgtcgcgcgaatatcgtcagcgaagcagcagcagcga	206
Db	9474	TTTHAKTTHRAASVSHSGSGKGNASHISGYHSHSRVAGSHSHSRASNAASHSGSHR	9533
Oy	207	cccccaacggctcaaacgcgttttgcggaagcggcgaagcgaagcggcgtttggaactgc	266
Db	9534	HAASRSRASAAGYSRASASVSTHNSRYSAAAGSTRRAASRAASGASHSHRSGSRYS	9593
Oy	267	cgcgttttgaagaacccgaagcagcagcagcagcagcagcagcagcagcagcagcagc	326
Db	9594	YSGHRSCTYSISRKAVGTAASRAASNMNTASASHVAGMTGYSAAAVAVASRVAGYSAA	9653

QY	327	acacgtgcacgaagcctttggcacaacacgaagcgctgtatcataacgcgcgcacatcgcg	386
Db	9654	NSRRERASRSAAVAVTHRRGASNSRKTENGHYHRYSGSTTHRRGGGYSGRARASHS	9713
QY	387	cagcgcagcattggcgagcgcctacatcaagcagaacgttc-ccgcgtacgcga	445
Db	9714	SKGNTHTHRRGTHRRHRRGVAVAAVSMTCGRMTSRASHSSRASSRPSRSRASNTRHCYSR	9773
QY	446	tgtacaacccgcgcgaataatcaagcagatagacaaaatcatctcagcgcgcgaggttcgcg	505
Db	9774	ASHYSGYHRRARVARTTRCNASVAAVYSTHRRGNMNSHVATHRHSHYSTGRASHASHS	9833
QY	506	gcaagggaaacacgcgcgcctaccagcatataagaaggttaaaacaatacataaagccctgc	565
Db	9834	VAAARHGSHVAASNSASHASHASHSRKRAKGYSRVAAHSAVSTHRRGNMNSNRGNICYSHC	9893
QY	566	gttcggcgcaagcaacacatcgtctcgtccgcagaccgcgtccctccctccctaagaagcgaggg	625
Db	9894	YSRRCYSSBASRSRVAVAAVNSRRAAGCYASNNAAASRYSRGGRSGNASHSYSSRG	9953
QY	626	aagcgcatgtgtgtatttcttcggcaaacccgcgtatataccatcagtcgtgcgcaaat	685
Db	9954	YYSVAGAGRGRAAVALASTYRAASSRRASRGNTSGYGRAGASHSKTYRYSASNGIMTITH	10013
QY	686	tg 687	
Db	10014	RG 10015	

RESULT	15
LOCUS	AC032019/c
DEFINITION	AC032019 167721 bp DNA linear HTG 04-MAY-2001
ACCESSION	Hom sapiens chromosome 17 clone RP11-277J6 map 17, WORKING DRAFT
VERSION	SEQUENCE, 21 unordered pieces.
KEYWORDS	AC032019 AC032019.3 GI:8081756
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT.
ORGANISM	human.
REFERENCE	Hom sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (bases 1 to 167721)
REFERENCE	2 (bases 1 to 167721)
AUTHORS	Unpublished
	2 (bases 1 to 167721)
	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
	Anderson, S., Balgutin, J., Barina, N., Bastien, V., Beda, F.,
	Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
	Campiano, A., Castle, A., Chospel, Y., Colangelo, M., Collins, S.,
	Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
	Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
	Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
	Grand-Pierre, N., Grant, G., Haas, B., Halford, A., Holtom, L.,
	Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
	Klein, J., Lacombe, K., Lamazeres, R., Landers, T., Lebecky, J.,
	Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
	McCarthy, M., McKean, P., McGurt, A., McKernan, K., McPheters, R.,
	Meldrum, J., Menkus, L., Minova, F., Miranda, C., Mieng, V., Morrow, J.,
	Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
	O'Neill, D., Olfart, T. M., Oliver, J., Peterson, C., Pierre, N.,
	Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
	Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
	Strange-Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
	Testeire, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
	Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
	Young, G., Zainoun, J., Zimmer, A. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome
COMMENT	Research, 320 Charles Street, Cambridge, MA 02141, USA
	On May 25, 2000 this sequence version replaced by:7705191.
	All repeats were identified using RepeatMasker:
	SmIt, A.F.A. & Green, P. (1996-1997)

AC023146 168414 bp DNA linear HTG 14-SEP-2000
 LOCUS Homo sapiens chromosome 17 clone RP11-648M17, WORKING DRAFT
 DEFINITION
 AC023146 SEQUENCE, 23 unordered pieces.
 AC023146
 AC023146.5 GI:10048096
 AC023146 HTG: HTGS_PHASE1, HTGS_DRAFT.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 168414)
 Waterston, R.H.
 The sequence of Homo sapiens clone
 2 (bases 1 to 168414)
 Waterston, R.H.
 Direct Submission
 Submitted (08-FEB-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Sep 10, 2000 this sequence version replaced gi:9392701.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0648M17
 ----- Summary Statistics -----
 Sequencing vector: M13, 60%
 Sequencing vector: plasmid, 40%
 Chemistry: Dye-primer ET; 60% of reads
 Chemistry: Dye-terminator Big Dye; 40% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 151435 bases at least Q40
 Consensus quality: 156494 bases at least Q30
 Consensus quality: 159176 bases at least Q20
 Insert size: 16800; agarose-fp
 Insert size: 166214; sum-of-contigs
 Quality coverage: 4.17 in Q20 bases; agarose-fp
 Quality coverage: 4.26 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 2233: contig of 2233 bp in length
 * 2234 2233: gap of unknown length
 * 5190 2233: contig of 2857 bp in length
 * 5191 5290: gap of unknown length
 * 5291 8044: contig of 2754 bp in length
 * 8045 8144: gap of unknown length
 * 8145 10817: contig of 2673 bp in length
 * 10818 10917: gap of unknown length
 * 10918 13707: contig of 2790 bp in length
 * 13708 13807: gap of unknown length
 * 13808 16819: contig of 3012 bp in length
 * 16820 16919: gap of unknown length
 * 16920 21212: contig of 4293 bp in length
 * 21213 21312: gap of unknown length
 * 21313 24508: contig of 3196 bp in length
 * 24509 28183: gap of unknown length
 * 28184 29283: gap of unknown length
 * 29284 33506: contig of 4223 bp in length
 * 33507 33607: gap of unknown length
 * 33607 39272: contig of 5666 bp in length
 * 39273 39372: gap of unknown length

FEATURES
 Source
 misc_feature
 39373 46456: contig of 7084 bp in length
 * 46457 46556: gap of unknown length
 * 46557 53287: contig of 6731 bp in length
 * 53288 53388: gap of unknown length
 * 53388 62153: contig of 8765 bp in length
 * 62153 62252: gap of unknown length
 * 62253 71803: contig of 9551 bp in length
 * 71803 71903: gap of unknown length
 * 71904 84711: contig of 12808 bp in length
 * 84712 84812: gap of unknown length
 * 84812 104254: contig of 19443 bp in length
 * 104255 104354: gap of unknown length
 * 104355 126746: contig of 22392 bp in length
 * 126747 126846: gap of unknown length
 * 126847 160548: contig of 33702 bp in length
 * 160549 160648: gap of unknown length
 * 160649 162273: contig of 1625 bp in length
 * 162274 162374: gap of unknown length
 * 162374 164324: contig of 1851 bp in length
 * 164325 164325: gap of unknown length
 * 164325 164325: gap of unknown length
 * 164325 166494: contig of 2170 bp in length
 * 166495 166594: gap of unknown length
 * 166595 168414: contig of 1820 bp in length.
 Location/Qualifiers
 1. 168414
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /clone="RP11-648M17"
 1..2233
 /note="assembly_name:Contig10"
 2334..5190
 /note="assembly_name:Contig11"
 5291..8044
 /note="assembly_name:Contig12"
 8145..10817
 /note="assembly_name:Contig13"
 10918..13707
 /note="assembly_name:Contig14"
 13808..16819
 /note="assembly_name:Contig15"
 16920..21212
 /note="assembly_name:Contig16"
 21313..24508
 /note="assembly_name:Contig17"
 24609..29183
 /note="assembly_name:Contig18"
 29284..33506
 /note="assembly_name:Contig19"
 33607..39272
 /note="assembly_name:Contig20"
 clone_end:SP6
 vector_side:left
 39373..46456
 /note="assembly_name:Contig21"
 46557..53287
 /note="assembly_name:Contig22"
 53388..62152
 /note="assembly_name:Contig23"
 62253..71803
 /note="assembly_name:Contig24"
 71904..84711
 /note="assembly_name:Contig25"
 84812..104254
 /note="assembly_name:Contig26"
 104355..126746
 /note="assembly_name:Contig27"
 126847..160548
 /note="assembly_name:Contig28"
 160649..162273
 /note="assembly_name:Contig3"
 162374..164224
 /note="assembly_name:Contig7"

/product="junctionp11n 3"
 /protein_id="AA140941.1"
 /db_xref="GI:1764245"
 /translation="MSSGGRFNFDDGGYCGGWEDKAGHGVCTGPKGGEYTGSM
 HGEVLGVGTPMSNTYOGTMAOGKRGIGLESKGVKWKYKGEWTHGFGRGVVEACAG
 NKAKEGWSNLODGYGTEYSDG"

BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others

ORIGIN

Query Match 4.5%; Score 40.2; DB 9; Length 125020;
 Best Local Similarity 10.3%; Pred. No. 13;
 Matches 58; Conservative 248; Mismatches 259; Indels 0; Gaps 0;

Qy 262 gcccccgcgttttcgagaacccggaagacataatgttcaagcggtacagc 321
 Db 51323 GACWTRKWCCTCMARMMARMAAMRAMAGARMMWMMWMMMBHAAAMARW 51264
 Qy 322 tgggaacatgtgcaagcgtttggacaacacgaagcgctctatcatcagccgac 381
 Db 51263 AAACKMWWSYRYKCMMSMMAAAMAMMMCTAAWYKKTMRMRSSKYSSTAYAWY 51204
 Qy 382 atcgacagctacgatttggcgagcgtacatcagccagcgttcctccgctgac 441
 Db 51203 ATMCWYCSMYAARMGTGSGKRRKRNASTSCMCCRRGCCMKKKKRWCTTTGMSG 51144
 Qy 442 gccatgtacaaccccgcaaaataaagcagatagacaaataatcagcagcgaggt 501
 Db 51143 GSKRMTTSMACWYKGGRRKTKMATSGCMRMGMRSGMRKYWCRCYKGMGRGRMS 51084
 Qy 502 cgcggcaaaagaaacccgctaccagcatcaggaagggtcaaaataatcatcaagc 561
 Db 51083 GYKGGSGVWMSMGWYSGRTSKYSTGTGCTSCSKMKKSKYSTSCCTKSYSCSMGRY 51024
 Qy 562 ctgcttgcggcggaagacacatcgtctcgtccgacacagtcctccctcaagaaagc 621
 Db 51023 SSYCCMGGRMKMGKSSCCASRSRGSCKSYKMAAGSWRCGRMGCMWYSSMKYSMAKY 50964
 Qy 622 ggggaagcgatgtgggtgattcttcgcaaacctgcttaccatcagcgtgagcga 681
 Db 50963 SSARBRGKMYKTKSCMKTYYSNHBKSSHDSGYSTRKSYBYWCMKMDSCWSDSHM 50904
 Qy 682 aaattggcacacgtcaaaagcgctgaaaacctgtttctgtcgtcgagcgtcgtgagc 741
 Db 50903 RDMWTKMKRYSSGCTRGYHHSAMCMKSMRDSGSGMYHMKSTYKSSASAKHMYSG 50844
 Qy 742 ggacaagtttcgatttgcacatccgcccgtccaaaggggaattgaacgagcacaagc 801
 Db 50843 MKSSVYBMDCMSYTSBSKSVBRMRMSGSYBRYRCKACVMMRYHRSTRSVMTGGRSC 50784
 Qy 802 catgtgcgcgcgtgttcaaccgca 826
 Db 50783 WVRDSVSRRRSVHTSMSSWRMCR 50759

RESULT 19
 AE005970/c 12680 bp DNA linear BCT 28-MAR-2001
 LOCUS Caulobacter crescentus section 296 of 359 of the complete genome.
 DEFINITION AE005970 AE0053673
 ACCESSION AE005970.1 GI:13424716
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Caulobacter crescentus.
 Caulobacter crescentus.
 Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 Caulobacter.

REFERENCE 1 (bases 1 to 12680)
 AUTHORS Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E.,
 Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R.,
 Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D.,
 Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
 Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
 Berry, K., Uitterlind, T., Tran, K., Wolf, A., Vamathevan, J.,

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and
 Fraser, C.M.
 Complete genome sequence of *Caulobacter crescentus*
 Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
 21173698
 2 (bases 1 to 12680)

1 (bases 1 to 12680)
 Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,
 Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I.,
 Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
 Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
 Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
 Berry, K., Uitterlind, T., Tran, K., Wolf, A., Vamathevan, J.,
 Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
 Fraser, C.M.
 Direct Submission
 Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 12680
 /organism="Caulobacter crescentus"
 /db_xref="taxon:69394"
 complement(156..1100)
 /gene="CC3067"
 complement(156..1100)
 /gene="CC3067"
 /note="Identified by match to protein family HMW"
 /codon_start=1
 /transl_table=11
 /product="peptidyl-prolyl cis-trans isomerase family
 protein"
 /protein_id="AAK25029.1"
 /db_xref="GI:13424717"
 /translation="MGVIGAVCALMIASGONKAYKPPRPDPNVAARVAVNGQVYAS
 DYKREAVAGLISECPDIDSEVRRQRIIDYDKLAAEAVKRLKIDPLDPAQRRLA
 AAKRRTLGMLVGVYKAVTDAIRKLAEDQKLSKSEETIRAOILVGSQAEESI
 KILATGASFDALEMRSDQVTRFNGDLGFTLDVMEAVAGALAKOAGALIGP
 AAEQWVIAKVBDRKTEPITLEARPQIIRFLTYDQVDRIDIELKRSAGAKYEMMIGKS
 OELPAGSAOEPASAPPELOGAPASAPAPAPAGAPAGAKQ"
 1425..4196
 /gene="CC3068"
 1425..4196
 /gene="CC3068"
 /note="Identified by match to TIGR protein family HMW
 TIGR00085"
 /codon_start=1
 /transl_table=11
 /product="preprotein translocase, SecA subunit"
 /protein_id="AAK25030.1"
 /db_xref="GI:13424718"
 /translation="MIGFAKKLFGSSNRRKVTATRYAKINAEAYALSDALKG
 KTAERKARLEKETLDDILNEAFVAVRERSKRVGLMRPDVOMGAVLHSGISSEMR
 TGEKTLVATLEPTYNALLEGGVHVTYNDVLARADAMQGVNRELGLSTGVYVNLG
 SOGEROAVRSDIYTGNNRERFDLRNLYSYSDENYORGNFAIVDEVDSTILIDEA
 RPPILISCPTEBRSSFTYIDVLVNLILUDSKSMEDHDKQKQVILITBEGQKIEIIM
 SGHILAEASAGLYDAANVYVHVNVNLAIRNLLKRYDVVDQKRAVEEEOEFMES
 TGRRLSEGLHQAIEAKEGADIQEPENOTLASYIONFYRLYKLSGMGTASTEORPD
 DIKMSVEIPTNRTIORIDDDDEQVYPTREKNEALIKQIDCHVRGQPLITGVNSTE
 KSEELSKLSTFSEFEKDKYKGIPIHOVLNRFHOEAVIYADAGYRGATITATNMG
 RETDIOLGSGTDMRFNRQORQRGEGLETITVEDDEERARLETILADKKAALAAAGL
 FVLGTEREHRIDQLRGRTGRGQDPERSFELSCEDDLIRIAGERLADAIKMPFV
 OEGEATIRKWLINALATQAQKREVEQENVEIRNLLKRYDVVDQKRAVEEEOEFMES
 DLSGLIHEMRVIDDLVLRHLPKAYAEQMDVGLTFRVYSGILDLPIAEWAEEG
 IADEMKERITKADEYVAOREVITTPROMSVKRSFLTQWIDLOMRHMHLDLNL
 VIGLKGVGORPLNRYKTEARSLEKILGDIRTYMTWTMLNVEIYAEHPYPTPLN
 LVEVHLIDLTGEMNAPAGITPEGLSTAOERALPVSLPEGMDRTRNAPCPGSGKRF
 KOCHRSIVR"
 complement(4297..5706)
 /gene="CC3069"
 complement(4297..5706)
 /gene="CC3069"
 /note="Identified by match to protein family HMW"
 /codon_start=1
 /transl_table=11

gene
 CDS

```

gene
CDS
    /product="major facilitator family transporter"
    /protein_id="AAK25031.1"
    /db_xref="GI:13424719"
    /translation="MSRPMIASSPAPRLGWYTAASFQEVYTDVTVVVALDAIG
    TPAAPPAOLMVDAYVLALAMLSAGIDGRPKRAFWGLFEGVAAGIAGLA
    GRLSTIVRAAGATGALLIPSLALINAAHGDRRRAVGMWTAGVIAAGPV
    LGLVETIGRMWFFVNVPTCLLAIITQVAPPTPPKPAAPDLNGOVALVAVT
    VGSALNFTYGMIVLGLYMLDMPGKAGLPLPLATFVSNLSGRVSLRGA
    RAVTAGVVAACGALTARLGAESGMLEMLPGFALIPGQGLAIPALTSLVSDR
    HEAGASGALNACROVGAAGVAFGALASVDAVAGVRSSGIASAMLLGAATVLP
    RGKSHPSSSAAEAQIHP"
    5779..6483
    /gene="CC3070"
    /complement(10059..12611)
    /note="identified by match to protein family HMM"
    /codon_start=1
    /transl_table=11
    /product="transcriptional regulator, Arsr family"
    /protein_id="AAK25032.1"
    /db_xref="GI:13424720"
    /translation="MDANIASPALIGDPTRAMTQVLODGRAPASALAMAGVTAQ
    AASNLAKLVDCGLAVERGRRHYRLASAEVAHAIEALSVIAPVSLIIPSPKA
    BALRDARCCYGLAGRLGYKCDALDRDLIRPADGRVATIPGCHAMEEDIGVSSQ
    LRGARGVAPPCLDMPERRRHLAGPLGVKLLAMTQGMALPENGRAVRLTNGARAL
    LERLSVLEDAQAAA"
    7362..7362
    /gene="CC3071"
    /complement(6508..7362)
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK25033.1"
    /db_xref="GI:13424721"
    /translation="MSYPSKVTLLSLAVLGAISGDKPPGNAKPPAPAPVAA
    PAATAPLEAPAPVPAPEIDGRKHEANLKLGOENRQOTLTLINGRGLMT
    DNDEPHILSGVTOVPANGVOTVETVTHYIGVADEPPAFDEGAGIESYDGL
    WRGAMFVGIQEOITYGVGVRSTELRPMSTPHKPKPKRCIPDAVSDTEKGY
    CLRPTGEGVELTDEKTDVFTFRVGNMRLELQAPKGVTDVNNPTVAYDEVIR
    GVPVSGDQ"
    7603..8514
    /gene="CC3072"
    /complement(7603..8514)
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK25034.1"
    /db_xref="GI:13424722"
    /translation="MLAVRMICGRPHFSRPGMLFQSRVMLALSLIAGLASCDC
    KLPGAAPTEBAKPVAPAPASAAPEPPNITAKDEBEIGAEELIKELGONNANG
    PTLTKNGANALVVDDEYLSLTNFWPLPDNGVQVYETVTTDLASGSPETTL
    YDRGVSEIETVPRGLMGAVLVTIGQDETMPPDDAARTLLDWSKPKRKEFSK
    CMTESVSDTEIKGYCVAPASOABNPMGEMNGDEVDVAVTFVGNAMVRELSPK
    AKLLNPDNPPTADETVKGPYSGNP"
    8558..9256
    /gene="CC3073"
    8538..9256
    /gene="CC3073"
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK25035.1"
    /db_xref="GI:13424723"
    /translation="MNEICRCGPRRSASPHCHGISGVADPGDFKRDGKTYTITOT
    TIGIKTITLIGAAATLMTAPPAHNADEKIGVYAGKASGKQGVAAVAGLLG
    AAGSNLAKNDGCGTAIGAVAGAGSLIGCKQKSDAEOVGATIGSGFRAATV
    QAPLVITEKMTVRSYVNLRAASTGERLGAVDSTFTQALGRTGDMGLVGGQG

```

```

gene
CDS
    /product="sensor histidine kinase/response regulator"
    /protein_id="AAK25037.1"
    /db_xref="GI:13424725"
    /translation="MRRVLILFVLVMTFGSGAMASDAFLAQRDLQSLGANEVR
    KPRDNDIDIERAGSALREAGQERLYGLMVLVAKSNQIQGRFDMATREVARRD
    KDATLEALVDLEVLAIRHESGQFADQKMDRELFKSGGDIRLMAGIEVRRLGIG
    RWAAARLAAELTGLDERRGHIAQPLAEIHOVSYSLSDQKEGALDHMAQALD
    ERDTEYVAKTERIYDIAFSAADIGELAAAEFEATLHKRLTMADQDPELVDFLCAR
    TAGIREAPERVITQCLAPAGILANPDRLOVRMIGORALALAOIGSGQAAARDLERLK
    TIPTPLASRDRPLEALVAYTIDYEGSADAFKRLDMWRKADGILSRNAYAKVAENS
    AALSELKVRKDESRLTEOVRLSRLASRVAIALLGLVYAGAAANVHOGASRR
    LRLAOREAASSTAKSAFLAWMSHELTPINGMIGLAQALRSQDLPQAEQVLDLD
    SGTLLVILNDLILDSKIEAGKLEIAPTAGDLVOTCARLVGYOPTAREGVRLSFLD
    DGBAPGMLMPFQVROCLSNLVNNAKFTSQGVYFALACYEAGDGRVRLRVADT
    GIGMAATVAKLFRFPPTQASTTRNKGGLGINTIRLVEMMRGDIQVSESEGVGS
    VFTLEWMDALALIEAPSENGATSESGAALQGRVVLVDDHPVRRVIRILPEPF
    VGVGFVSSAYVYRA"
    9489..10010
    /gene="CC3074"
    /complement(9489..10010)
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAK25036.1"
    /db_xref="GI:13424724"
    /translation="MAAOSTOMTQRRLITSRLPLPBARCMLTCTWALSALTATMAIG
    FTGGLYHLPFPODLALMLVTLIPALGEIEFRAVLTPGEGRRPVEIGVAFGL
    VYLMHVVEALTFPLPNNAPVFLRDFLCCAVIGLGAATKRVTSVPAVLLMALV
    VNOTWLGVSALS"
    10059..12611
    /gene="CC3075"
    /complement(10059..12611)
    /note="identified by match to protein family HMM"
    /codon_start=1
    /transl_table=11
    /product="transcriptional regulator, Arsr family"
    /protein_id="AAK25037.1"
    /db_xref="GI:13424725"
    /translation="MRRVLILFVLVMTFGSGAMASDAFLAQRDLQSLGANEVR
    KPRDNDIDIERAGSALREAGQERLYGLMVLVAKSNQIQGRFDMATREVARRD
    KDATLEALVDLEVLAIRHESGQFADQKMDRELFKSGGDIRLMAGIEVRRLGIG
    RWAAARLAAELTGLDERRGHIAQPLAEIHOVSYSLSDQKEGALDHMAQALD
    ERDTEYVAKTERIYDIAFSAADIGELAAAEFEATLHKRLTMADQDPELVDFLCAR
    TAGIREAPERVITQCLAPAGILANPDRLOVRMIGORALALAOIGSGQAAARDLERLK
    TIPTPLASRDRPLEALVAYTIDYEGSADAFKRLDMWRKADGILSRNAYAKVAENS
    AALSELKVRKDESRLTEOVRLSRLASRVAIALLGLVYAGAAANVHOGASRR
    LRLAOREAASSTAKSAFLAWMSHELTPINGMIGLAQALRSQDLPQAEQVLDLD
    SGTLLVILNDLILDSKIEAGKLEIAPTAGDLVOTCARLVGYOPTAREGVRLSFLD
    DGBAPGMLMPFQVROCLSNLVNNAKFTSQGVYFALACYEAGDGRVRLRVADT
    GIGMAATVAKLFRFPPTQASTTRNKGGLGINTIRLVEMMRGDIQVSESEGVGS
    VFTLEWMDALALIEAPSENGATSESGAALQGRVVLVDDHPVRRVIRILPEPF

```

```

Query Match
Best local similarity 50.3%; Pred. No. 32;
Matches 95; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 555 caaagccctggcttgggggaagcaaccatctctgcccgaacgccccctccca 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11151 CAGGCGGTGGCCCTCGATCACCTTACCCCGGAGCGGCGATCGATCTCT 11092

QY 615 aagaagcggaagcgtaagtgatcttcgcaaacctgctataacatgaagct 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11091 CGACAGCGGCGACACCTCTGCTGCTGTAACGACATTTGACCTTCCAGATCGA 11032

QY 675 ggcgcgaataatggcaacgctcaaaagcgctgaacacctgtttctcgtggaagcgct 734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11031 GCGGGGGAAGCTTGAGATGCGCCGCGGCGGTGATGTTCAAGACTGCGCGCCT 10972

QY 735 gacctgacg 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10971 GGTGGCGCG 10963

```

```

RESULT 20
AL591675/c 166919 bp DNA linear HTG 06-JUN-2001
LOCUS Mmus musculus chromosome 2 clone RP23-420L2, *** SEQUENCING IN
DEFINITION Mmus musculus chromosome 2 clone RP23-420L2, *** SEQUENCING IN
ACCESSION AL591675
VERSION AL591675.5 GI:14330189
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 166919)

```


QY 443 ccatgtacaaaccgcggaatcacaagcgatagacaaatcatgcagcgcgcgaggttc 502

Dr J. C. Sanger, Sanger Centre, Hinxton, Cambridgeshire,
Submitted (17-OCT-2001)
C10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequest@sanger.ac.uk
On Oct 21, 2001 this sequence version replaced g1:16151503.

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
Project Information
Center project name: BM34512
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 184803 bases at least Q40
Consensus quality: 185238 bases at least Q30
Consensus quality: 185716 bases at least Q20
Insert size: 186434; sum-of-contigs
Insert size: 181929; 2.7% error; agarose-fp
Quality coverage: 11.78x in Q20 bases; sum-of-contigs
Quality coverage: 12.11x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

FEATURES	Location/Qualifiers
source	1. .187334

misc_feature

misc_feature

misc_feature

misc_feature

Center, Millenium Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="RP11-10624"
		/clone_id="RPCI-11 Human Male BAC"
		1..2044
misc_feature		/note="assembly-fragment"
		2145..4434
misc_feature		/note="assembly-fragment"
		4535..14152
misc_feature		/note="assembly-fragment"
		14253..30085
misc_feature		/note="assembly-fragment"
		30186..44736
misc_feature		/note="assembly-fragment"
		44837..60580
misc_feature		/note="assembly-fragment"
		60681..77501
misc_feature		/note="assembly-fragment"
		clone_end:SP6
		vector_side:left"
		77602..102683
misc_feature		/note="assembly-fragment"
		102784..138039
misc_feature		/note="assembly-fragment"
		23838 c 23904 g 43307 t
BASE COUNT	46183 a	807 others
ORIGIN		

Query Match	4.2%	Score 38;	DB 2;	Length 138039;
Best Local Similarity	52.5%;	Pred. No. 51;		
Matches	83;	Conservative	0;	Mismatches 75; Indels 0; Gaps 0;

BASE COUNT	ORIGIN
1086 a	822 c
841 g	899 t

2072 GACCACTGAGATCGGTGCCTCAGACCAACCGTAGTGAGGAGAACATTGCCAAGGCTCTG 213

[illegible]

RESULT	29
AC017294/c	LOCUS
AC017294	DEFINITION
AC017294	63904 bp DNA linear HTG 09-DEC-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***	in ordered
AC017294	PIECS
AC017294	AC017294.1 GI:5553692
HTG; HTGS_PHASE2.	fruit fly.
SOURCE	Drosophila melanogaster
ORGANISM	Drosophila melanogaster

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 63904)	Adams, M. and Venter, J. C.	Direct Submission	
		Submitted (09-DEC-1999)	Celera Genomics, 45 West Gude Drive, Rockville, MD, USA	This sequence was identified as CDM:10210186 by the submitter. For more information on this record e-mail to fly@celera.com
				* NOTE: This is a 'working draft' sequence.
				* This sequence will be replaced
				* by the finished sequence as soon as it is available and
				* the accession number will be preserved.
				Location/Qualifiers

FEATURES	location/qualifiers
source	1. 63904 /organism="Drosophila melanogaster" /db_xref="taxon:7227"
BASE COUNT	17101 a 14155 c 14516 g 18132 t
ORIGIN	

Query Match	4.1%	Score 37	DB 2	Length 63904
Best Local Similarity	47.8%	Pred. No. 89		
Matches 139; Conservative	0	Mismatches 150	Indels 2	Gaps 1

```
Qy      167 gcaatcgtcgccaatatgctgcagcgaagcatgaaaccccccacaacgctcaaacgccg    226
          |  |||  |  |||||  |||  |  |||  |||  |  |||  |||  |  |||  |||  |
Db     23100 GTAATCTTGCGCCATATACGATGCATAAATTACTTCGGAAATTCTTAAAGGAATCAATCCAC    23041
```

QY 227 ttlttgc--ggaacgcgcacaagaagcgttttgaaactgtcccccggcttlttcagaaac 284
||| ||| ||| | ||| ||| ||| |||
Db 23040 TTTCGTTCCAGGAATCTCGAAGAGCTCTTCCGTGCTCAACTCGGCCAATGGATATACC 22981

QY 225 ggaagacatagaacaacatcttcaagcgyttacaagcgtggaacatgtgcagcagcttc 344
| | | | | | | | | | | | | | |
Db 22980 GCCTGATCTAACGCCCACTGTGAATCACCGCTCATCCGGAGCGGTTGCTGCACGACGA 22922

QY 345 ggacaaacagcaaggcgctgtatcattcatccgcccgcacatcgacgctaagaattggcg 404
||| ||| | | | | | | | | |
Db 22920 GCCCACACAAACGGAGATGAAGACGATCCAGGAGATCCTGTGC GCTTCCGCCAGATCAC 22861

QY 405 agctacatcagccagcagcttcgttcccgcgtgaccgcgatgtataaac 455
||| ||| ||| ||| | ||| ||| |||
Db 22860 ACCCGACGGCAGCGAGGTGGCTGCATGAAGGCCATGCCCCCTGTTCCACC 22810

RESULT	30	
AC092239		
LOCUS	AC092239	
	156641 bp	PMN
	11-00000	
	TRF 50	TRN 00000

DEFINITION	Drosophila melanogaster, chromosome 2L, region 25F-26X, BAC clone 350-00N-2000
ACCESSION	AC092239.1
VERSION	GI:14578120
KEYWORDS	fruit fly.
SOURCE	HTG.

ORGANISM *Drosophila melanogaster*
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephyridiidae; Drosophilidae; Drosophila.
1 (bases 1 to 166441)
REFERENCE
Celniker, S.E., Adams, M.D., Krommler, B., Tyler, D., Wan, K.H.,

Holt, R. A., Evans, C. A., Gocayne, J. D., Amantides, P. G., Brandon, R. C., Rogers, Y., An, H., Baldwin, D., Bonzon, J., Beeson, K. Y., Busan, D. A., Carlson, J. W., Center, A., Champe, M., Davenport, L. B., Dietz, S. M., Dodson, K., Dorsett, V., Doup, L. E., Doyle, C. S., Dressnek, D., Fafian, D., Ferreira, S., Frisoe, E., Galle, R. F., Gary, N. S., George, R. A., Gonzalez, M., Houch, J., Hoskins, R. A., Hostin, D., Howland, T. J., Ilegwam, C., Jallali, M., Kuse, D., Li, P., Mattel, B., Moshfegh, A., McIntosh, T. C., Moy, M., Murphy, B., Nelson, C., Nelson, K. A., Nunoo, J., Paclel, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Plounavong, S., Pittman, G. S., Puti, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskis, R., Tector, C., Williams, S. M., Zaveri, J. S., Smith, H. O., Rubin, G. M., and Venter, J. C.

Sequencing of Drosophila chromosome 2L, region 25f-26x

TITLE
JOURNAL

REFERENCE
AUTHORS

2 (bases 1 to 166441)
Celniker, S. E., Adams, M. D., Krommiller, B., Tyler, D., Wan, K. H.,
Holt, R. A., Evans, C. A., Coccyne, J. D., Amanatides, P. G., Branton, R. C.,
Rogers, Y., An, H., Baldwin, D., Bazzon, J., Beeson, K. Y., Busam, D. A.,
Carlson, J. W., Center, A., Champe, M., Davenport, L. B., Dietz, S. M.,
Dodson, K., Dorsett, V., Doup, L. E., Doyle, C., Dresnak, D., Farfan, D.,
Ferreira, S., Frisbe, E., Galle, R. F., Garg, N. S., George, R. A.,
Gonzalez, C. M., Houck, J., Hoskins, R. A., Hostin, D., Howland, T. J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Mosheiff, A.,
McIntosh, T. C., Moy, M., Murphy, B., Nelson, C., Nelson, K. A., Nunoo, J.,
Pachlebb, A., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G. S., Puri, V., Richards, S., Scheeler, F.,
Stavert, M., Strong, R., Svirskas, R., Tector, C., Williams, S. M.,
Zaveril, J. S., Smith, H. O., Rubin, G. M. and Venter, J. C.

TITLE
JOURNAL

Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,

LOCUS RME03647 189333 bp DNA linear BCT 16-AUG-2001
DEFINITION Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete psymb;
segment 6/6.
ACCESSION AL603647 AL591985
VERSION AL603647.1 GI:15141286
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE
1 (bases 1 to 189333)
Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P.,
Vorholter, F.J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J.,
Goulding, B., and Puhler, A.
From the Cover: The complete sequence of the 1,663-kb psymb
megaplasmid of the N2-fixing endosymbiont Sinorhizobium meliloti
Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9889-9894 (2001)
11481431
PUBMED
epub ahead of print
2 (bases 1 to 189333)
Weidner, S.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld,
Biologie IV (Genetik) Universitaetsstr 25, D-33615 Bielefeld,
Germany
Submitted on behalf of Universitaet Bielefeld, Biologie IV
(Genetik) Universitaetsstr 25, D-33615 Bielefeld, Germany and
Department of Biology, McMaster University, 1280 Main Street West,
Hamilton, Ontario, L8S 4K1 Canada
Mailto:Stefan.Weidner@genetik.uni-Bielefeld.DE
pexo, psymb.
FEATURES
source
Location/Qualifiers
join(1..63290,join(64592..78847,join(80072..80112,
join(80617..82500,join(83050..162434,163659..189333))))
/organism="Sinorhizobium meliloti"
/plasmid="psymb"
/strain="1021"
/db_xref="taxon:382"
/locus
complement(126..932)
/gene="idnol OR Smb20692"
complement(126..932)
/gene="idnol OR Smb20692"
/EC_number="1.1.1.69"
/function="Small Molecule Metabolism; Degradation; carbon
compounds"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="putative 5-keto-D-gluconate 5-reductase protein"
/protein_id="CAC49799.1"
/db_xref="GI:15141287"
/translation="MMDGASNAAPRTGRLPAFLDSGRVALLTGSSGIGITMAALC
EAGSYIINGRDEARLARGARALEHRCFTVGSAPDVATKSAIGRAVADILAEERID
ILVNNAGIOHRTPLHEFPDARFVETINTSATLVQAVQGMTEREGTIIINCSV
QSELAPSTAPYAAASKGGLKMLTKGMALDMRGYIGRVNGLAPGYFKTELSNLSVDEK
FSTWLEQRTPLGRMDGTGLAAAVFLASASSFVTGHILYVDGITSCL"
complement(1063..1361)
/gene="Smb20693"
complement(<1063..1361)
/gene="Smb20693". 1361)
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49800.1"
/db_xref="GI:15141288"
gene
CDS
/translation="MPDYASKGVGQOLAKAFATNGOPGTFRSMPLPGYATIDSTEA
SRNDPSRASLILARIIPRAAGAPRTILPARRSIPRRHPIITSARASSPWTAGQWA"
1508..1684
/gene="Smb20694"
1508..1684
/gene="Smb20694"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49801.1"
/db_xref="GI:15141289"
/translation="MGADDDPMAALDARVRVIERLRYADMSAVPDINAGTNPAT
MLGNRCAGFILGTA"
complement(1765..2271)
/gene="Smb20695"
complement(1765..2271)
/gene="Smb20695". 2271)
/function="MISCELLANEOUS; Unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49802.1"
/db_xref="GI:15141290"
/translation="MTLLPVALGTELVAMRLDOAVHAPTWDSGEGAPKSGRNSNG
VRAVCSIDPATALLLEVAVHKGFRALDVAHTRAADAGDYVYVDKSPNPMA
RPGICAGQOAGDILLRRHRAVIPSAPSHSNMVLFGAAPPAVALKQEPFALD
TRLHPPT"
complement(2268..2753)
/gene="Smb20696"
complement(2268..2753)
/gene="Smb20696"
/function="MISCELLANEOUS; Unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49803.1"
/db_xref="GI:15141291"
/translation="WALAARPTPAAGGGLQKLETLIGSRVLAHHLTNALDAHELL
LRGLPASALDHLVGNLVETIGKTESLEKAVGMSLRWRRKETPKRPLSOSGRAMKE
AEILVAKTIDIFGSOAEAEQWLERPAVNGIDGRRPIDLGTTPGVELVDHLDRLEYGY
A"
complement(2855..4549)
/gene="Smb20697". 4549)
complement(2855..4549)
/gene="Smb20697"
/function="MACROMOLECULE METABOLISM; Macromolecule
degradation; degradation of proteins, peptides,
glycopptides"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="putative peptidase, similar to arginine
utilization protein RcoB"
/protein_id="CAC49804.1"
/db_xref="GI:15141292"
/translation="MTNAQDETVSRAGSRIDSBRVQFALRMTSMPSGSETGPEASFA
DRHLGLGELPYFRHPODGLASHEPELTRVAVLVREGTKRTIVMAGHPVTSTD
NYHRLKALGDSPLAKDALIESLSARGRSEDEPRALODASGDTPLPGRLDMKSGL
AVAIACIEGFAADTDRGNMLVATDEBERESGMSLRDLGVLKRDVETIAGIN
LDVTSDGDSSEGRAYAGTIGKLLPVALVIGSSHASYPPEEGVSAQMAAGILAELE
GNASLADRDNDISPPICLEARLDKRGYEVTPPERFWIAFWMLYHSMWADALFERFR

```

EEVLGTGANEAIETFAAQAQAEYKGLVRSAGSLPAKPRLLSFOELRAAARVGHGFDA
FYAEKEREFAQSDNPVLAIRQLTEWLVLGIRASGPATVIGSGHLPSPHRLAEND
RSLHOAIKARAGILGMPKREPHOKFERVHEPAPFVLELVSEIARTFLGDDRHD"
complement(4701..6353)
/gene="SMB20698"
/complement(4701..6353)
/function="CELL PROCESSES; Transport of small molecules"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein, similar to E. coli
multidrug resistance protein EmrB"
/protein_id="CAC49805.1"
/db_xref="GI:15141293"
/translation="MTIATPSPRISNAGGILVAGIVATITLTAIASVNLGRGDI
IGDTVATPDEFAMLDGTTTFLKIGPMASWMLSRDPNNLVGSLTMAGCIIAM
TARLDLVALRMIOGFSGGTLVGOAILTFPPSHROPLDAFFPMGSVAPATIP
ALOGILDSQWTFIFSIPLAAAGILLADGPVARRAPRPDMVIGSLISAL
FCETVYESQSRMDPEERILMLAVIGATILAIIGQVILAKGGLDFLFEDEF
CPAFIVFAVAGALFSGALIPSPAVSVIATPTDAGOLLPSGALGALLIAFLM
QLRVVPVATVPGLIMMAAMMLSGSGAGDMAAILIRGAGGFLSTLI
ASNLMSRLNASTIGLFTNGROLGLIGISALQTLIENHVSNIIVANVITAGAPAV
ADRLTTTALLTRAKGMAAASRGAASLGRVAVAGSTVIAFDTPAIGLFLVIAAP
VIVGIRIGLARAKARAEKRGAVAPSGHHTLPKRPAS"
complement(6350..7447)
/gene="SMB20699"
/complement(6350..7447)
/gene="SMB20699"
/function="CELL PROCESSES; Transport of large molecules;
protein, peptide secretion"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative protein secretion protein, HlyD family,
similar to E. coli EmrK"
/protein_id="CAC49806.1"
/db_xref="GI:15141294"
/translation="MESLIEKVAIITGARSNGVAVARPNKKMIGIALCGLAVVVV
AGMTWARESGAASDNAVYRGVDTSLAPKAVGATVAVEDNCAVAGVLFRIIDRD
VRAKIQAQVAVARLITNVDAEMALHAIROAEOHRSVVALNIAAAYDRR
ELIRSETISQAHDESAPAKSRAEAVVAVASATVEAOORIVLAQOREAVAAVNOA
EAARDLAGIDLESTVAVAPVGVIGNRQVAVGRLVAPGASLIDIVFLDVMVITANKE
TOLEHIRPGORASTITIDYKSGALEGVDSFAPGSGASFLPADNATGNEVAVOVOR
PVKIRFAGNPLSGRLIPGISAARVEIDLEGGs"
complement(7534..7947)
/gene="SMB20700"
/complement(7534..7947)
/gene="SMB20700"
Query Match 4.1%; Score 37; DB 1; Length 189333;
Best Local Similarity 48.8%; Pred. No. 94;
Matches 100; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 422 agcttcgttccgcgtgcacatgtcaaacgcgcgaataaagcgatagacaaa 481
DB 112504 ATCTCCCGACACCTTACCGCACCGTATTCTGCTCGCGTCATGAAAGGCTCGACACG 112563
QY 482 tcatgcagcgagcgaggttcgcgcgaagaanaacgcgcctaccgcacatacaagg 541
DB 112564 AAGAGGCGGCGCCCTGCTGCTCAAAACGGAAGGCAACGCAAGGCTGCACCGGG 112623
QY 542 tcaacaacaatcaacaagccctgcgttcgcgcgaagaacacatcgctccgcgcgaacaa 601
DB 112624 CGCGCGCATCTCTGAAAGAGCGCTGAGCGGCGGCGGCTCTCTCGATGCTT 112683
QY 602 tccctccctcaagaagaagcgagga 626

```

DB 112684 TTCCTTTCGCGGGAAGGCTTGCGA 112708

RESULT 33
AE003611
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 1420001386055 section 4
of 63 complete sequence.
ACCESSION
AE003611 AE002690
VERSION
AE003611.2 GI:10728594
KEYWORDS
HTG.
SOURCE
ORGANISM
fruit fly
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.

REFERENCE

ADAMS, M.D., CELINKER, S.E., HOLT, R.A., EVANS, C.A., GOCAYNE, J.D.,
1 (bases 1 to 259973)
/db_xref="taxon:7227"

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

Location/Qualifiers
1..259973
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
join(16590..16757,16815..17687,17749..>17886)
/gene="CG7236"
/product="CT22313"
/db_xref="FLYBASE:FBan0007236"


```

Db 241198 GCTGATCTAACCGCCATCTGAAATCACCCTCATCCGAGACGGGTCTGCGAGGACA 241257
Qy 345 ggacaaacaggaagggtgtatctatcaccgcacatcgagctacgattggcg 404
Db 241258 GGGCCACCAACAGGAGATTGAAGAGATCCAGGAGATCTCTGCGCCCTCCGCAAGATCAC 241317
Qy 405 acgctacatcagcagcagcttcctccgtcgcgtgacgcagcatgaacac 455
Db 241318 ACCCGACGACGAGGTGGCTGATGAGGACCATCGCCCTGTTCGACAC 241368

RESULT 34
AP001517 294250 bp DNA linear BCT 10-JAN-2001
LOCUS Bacillus halodurans genomic DNA, section 11/14.
DEFINITION AP001517 BA000004
ACCESSION AP001517.1 GI:10175500
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacillus halodurans DNA.
Bacillus halodurans
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
1 (sites)
Takami, H. and Horikoshi, K.
Reidentification of facultatively alkaliphilic Bacillus sp. C-125
to Bacillus halodurans
Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
2 (sites)
Takami, H., Takaki, Y., Nakasone, K., Sakiyama, T., Maeno, G.,
Sasaki, R., Hirama, C., Fujii, F., and Masui, N.
Genetic analysis of the chromosome of alkaliphilic Bacillus
halodurans C-125
Extremophiles 3 (3), 227-233 (1999)
99411980
3 (sites)
Takami, H., Nakasone, K., Hirama, C., Takaki, Y., Masui, N., Fujii, F.,
Nakamura, Y., and Inoue, A.
An improved physical and genetic map of the genome of alkaliphilic
Bacillus sp. C-125
Extremophiles 3 (1), 21-28 (1999)
99184645
4 (sites)
Takami, H., Masui, N., Nakasone, K., and Horikoshi, K.
Replication origin region of the chromosome of alkaliphilic
Bacillus halodurans C-125
Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
99356711
5 (sites)
Takami, H., Takaki, Y., Nakasone, K., Hirama, C., Inoue, A. and
Horikoshi, K.
Sequence analysis of a 32-kb region including the major ribosomal
protein gene clusters from alkaliphilic Bacillus sp. strain C-125
Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
99209008
6 (sites)
Takami, H.
Genome analysis of facultatively alkaliphilic Bacillus halodurans
C-125
(in) Extremophiles in deep-sea environments (Ed.):
HORIKOSHI, K. TSUTSUMI,
: 249-284; Springer-Verlag (1999)
7 (sites)
Takami, H., Nakasone, K., Ogasawara, N., Hirama, C., Nakamura, Y.,
Masui, N., Fujii, F., Takaki, Y., Inoue, A. and Horikoshi, K.
Sequencing of three lambda clones from the genome of alkaliphilic
Bacillus sp. strain C-125
Extremophiles 3 (1), 29-34 (1999)
99184646
8 (sites)
Takami, H. and Horikoshi, K.
Analysis of the genome of an alkaliphilic Bacillus strain from an
Industrial point of view

JOURNAL MEDLINE 20263314
REFERENCE 9 (sites)
AUTHORS Nakasone, K., Masui, N., Takaki, Y., Sasaki, R., Maeno, G., Sakiyama, T.,
Hirama, C., Fujii, F. and Takami, H.
TITLE Characterization and comparative study of the rrm operons of
alkaliphilic Bacillus halodurans C-125
JOURNAL MEDLINE 20426005
REFERENCE 10 (sites)
AUTHORS Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N.,
Fujii, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and
Horikoshi, K.
TITLE Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis
Nucleic Acids Res. 28 (21), 4317-4331 (2000)
20512582
11 (bases 1 to 294250)
AUTHORS Takami, H. and Takaki, Y.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Hideyo Takami, Japan Marine Science and
Technology Center, Deep-sea Microorganisms Research Group; 2-15
Natsushima, Yokosuka, Kanagawa 237-0061, Japan
(E-mail: takami@jamstec.go.jp,
URL: http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html,
Tel: 81-468-67-3895, Fax: 81-468-66-6364)
FEATURES
source
Location/Qualifiers
1..294250
/organism="Bacillus halodurans"
/db_xref="taxon:8665"
/note="alkaliphile"
236..823
/gene="BH2879"
236..823
/gene="BH2879"
/codon_start=1
/transl_table=1
/product="spore cortex lytic enzyme"
/protein_id="BAB06598.1"
/db_xref="GI:10175501"
/translation="MKKVLILFSCITLITSTLTNKAESKHOVSGDPLLYLSEQY
GVPMALIRINERSNTTYRGEOLITPARVYASRDLRLRYHAEGEPYAGVAVAA
VWLVLRVDPSPFPDVTVDVINEVSGTYAESVONGRIORPANDAKRAYADEALQFNG
QNGSLYFNPNTAPSTPYSTROQIVIGNHIFAK"
955..1605
/gene="BH2880"
955..1605
/gene="BH2880"
/note="BH2880
unknown"
/codon_start=1
/transl_table=1
/protein_id="BAB06599.1"
/db_xref="GI:10175502"
/translation="MRRISRTSSTRTGYRATVNPPLGTSVSRVVEPPPIETANET
AKPSNPFVLSYERYKQVFEKELVHTLQALIGTNKELISQCTLIC
KYNEFLKSIRKIDOLFQSYMDIOALHIOKELEAGLINDNGTLTNSDVFRQ
LSTSQDPMAYIKRKSLILNQYHTISQLVQPNHQMSPYPPGPELKGFIIEKG"
1694..2977
/gene="BH2881"
1694..2977
/gene="BH2881"
/note="BH2881
unknown conserved protein in others"
/codon_start=1
/transl_table=1
/protein_id="BAB06600.1"
/db_xref="GI:10175503"
/translation="MSEHQSYSNDNLIDLNHTLTNDHLQTLKLGEMAKRLMNL
FLVGGVDMLEKGVPGDIDYTEEDALAFSGNVANVYLGKVKHHPATATWGAEN
IKLIDIVSARAEVYKPGALPTIRSHITTDLARPSFINAAHILHPASTYQOLDFPH
GRHDLTNGILRIHLSQSFIDPTRLRGVRSRNPYEQKTANLALATLALATPALNLE
ANVSPRIYHEIKLCHETDPVSPSKLEDLHWQALGLTFSSSSATHLRLQDEON

```

Query Match	4.1%;	Score 36.8;	DB 1;	Length 294250;
Best Local Similarity	51.2%;	Pred. NO. 1.1e+02;		
Matches 86;	Conservative 0;	Mismatches 82;	Indels 0;	Gaps 0;

```

gene
92..508
/ gene="mcm2"
/ note="VNG0673G"
CDS
92..508
/ gene="mcm2"
/ note="mcm2"
/ codon_start=1
/ transl_table=11
/ product="methylmalonyl-CoA mutase"
/ protein_id="AAG19166.1"
/ db_xref="GI:10580260"
/ translation="MSESTRNRCIAKVTGLGDHGRGAHYITRAFRNGEVIYSGI
HKRPDIVOAAVOEDVDVLTGISGHNLTLPKTIIEGLAEYDAEDLTLYIVGGITPD
EDPADLEAGVCYAIRGFTPMADITSTIESNAHRD"
510..1484
/ gene="VNG0674C"
/ CDS
510..1484
/ gene="VNG0674C"

```


VERSION	AL162756 AL157959
ACCESSION	AL162756.2 GI:7380091
KEYWORDS	
SOURCE	
ORGANISM	Neisseria meningitidis Z2491. Neisseria meningitidis Z2491. Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE	1 (bases 1 to 329861) Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T., Davies,R.M., Davis,P., Devlin,K., Feltham,T., Hamlin,N., Holtzoyd,S., Jagsels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Rance,M.A., Rutherford,K.M., Simmonds,M., Skellern,J., Whitehead,S., Spratt,B.G. and Barrell,B.G. Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
TITLE	Nature 404 (6777), 502-506 (2000)
JOURNAL	20222556
MEDLINE	2 (bases 1 to 329861)
REFERENCE	Parkhill,J.
AUTHORS	Direct Submission
TITLE	Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
JOURNAL	sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT	Notes: Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/). location/Qualifiers
FEATURES	1..329861
SOURCE	Feature(s) described in Neisseria meningitidis Z2491"

Query Match	4.1%	Score 36.4	DB 1	Length 10194
Best Local Similarity	44.3%	Pred. No. 1.2e+02		
Matches 148	Conservative	0	Mismatches 186	Indels 0
Gaps				0

QY	477	caaaatcatgcaagcgcgaggttcgcgycgcaagaagaaaaacgcgcctaccagcataca	536
Db	6052	CAAACTTTTCGACTACGCGCAACGGGGTAACCGTGGGCTTCAACCTGGCCAGCAAAACCCA	6111
QY	537	agggtgtcaaaccaatcatataaagccctgcgttcgcgycgaagaacacatcgttcctgcgcga	596
Db	6112	AGGGAGCAAAAGACATCATATAAATCAAAAGCGTGTGCTTTGGACGACACAAGCGCCGACCG	6171
QY	597	ccacgtccctcccccccaagaagcgcgaggaugcgtaiggttgatctcttcgycgaacc	656
Db	6172	CCTGGCCCTGTGGCCCCCGGAACGGGTGTCACACACATTCACAAATTTCAAAAGTGTGTCGA	6231
QY	657	tgccataaccatgacgcgtgcgycgcaaaatttgcacacgtaacgaagcgatgaaacctglt	716
Db	6232	GAACGGGCATTTGAACCTGGCCGACGCAAAATCCGCCAAGTGTTCGCGTCCGAAACAGAA	6291
QY	717	ttctgtgtggaagcgcctgcgttcgycgagaaagtgttgatttgacgttcgcgccgcgtca	776
Db	6292	TTTGCGCGGCGACGGGAGCGGGTCAAAAAGCCGGTTTATGTATTAAGCAACAAACGGGCA	6351
QY	777	aggggaattgaacgcgcgaacaaagcccatgatgcc	810
Db	6352	GACCGGCGCTGAATGCCACTACGCAAAAAAAC	6385

RESULT 38	NMA522491	329861 bp	DNA	linear	BCT 04-DEC-2006
LOCUS	NMA522491				
DEFINITION	Neisseria meningitidis serogroup A strain Z2491 complete genome; segment 5//.				

```

/misc_feature
/transl_table=11
/product="putative threonine synthase"
/protein_id="CAB84676.1"
/db_xref="GI:7380093"
/db_xref="SPTREMBL:Q9JU91"
/translation="MKYISTRGETAHAKPFSEVLMLMGAPDGLMLPEHYOIGREALD
KMRGLAYELAEIMRLFTVDIPEDDLRLDIINRYTEAAGTKEITPYRLTSDSIRKIO
ALSNGLTAPLAEIMRLFTVDIPEDDLRLDIINRYTEAAGTKEITPYRLTSDSIRKIO
VEMLSPPDGKMSAFORAMYSLODGNININIAVKGMFDCCODIYKAVONDAAREKMHG
TVNSINMGRIYAQVYVYFAGYFKAKNSDEOVSCVPSGNDLCAHIAKOKLPYR
RLIVATNENDYLDDEFRTGAYRPRNSATITVYSSMDISANSNVEFEDLMRDPO
EINTIMAEVAAKGFNLFDKVGKGYFTSGSTADRLATIRQYEDOEQELIDPH
TADGVKVAEREEREYVLCLETAALAFATIRAVODVAIPRPALEGHLENPORV
QTPNSADAVKGIIEOTLA"
75.1780
/gene="hcr"
/feature="pfam match to entry PF00291 S_T-dehydratase,
pyridoxal-phosphate dependent enzymes, score 231.80,
E-value 1e-65"
/complement(922..931)
/label=DUS
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
2094..2888
/gene="NMA1441"
2094..2888
/gene="NMA1441"
/feature="NMA1441, len: 264 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1441"
/protein_id="CAB84677.1"
/db_xref="GI:7380094"
/db_xref="SPTREMBL:Q9JU90"
/translation="MYVMOETHPSTLLDELAQKOAIAIPHLADGTXYMKRAGSH
NARKRYALIGVARYKLGVLKVPSPGEGPATETSKRLXELRAAIAINPELLAVK
NALMGNLEGIPLDPTORQAEKAKALAMAGALAIRHKKQDFLSQAFARNMSDG
KNISFLPDEDDSEVLTIAQCQARDWLCYHSTLILKNGLEAAAEKNGVLSIDP
AEIKLTAIVYKPLIRLEPRMGDARLAAISILSIADMP"
2904..2937
/feature="Stem loop containing DNA uptake sequences: acgat
gccgtctgaa gccgtctgaa atata"
2909..2918
/feature="Stem loop containing DNA uptake sequences: acgat
gccgtctgaa gccgtctgaa atata"
2909..2918
/label=DUS
/feature="Core DNA uptake sequence: gccgtctgaa"
/complement(2923..2932)
/label=DUS
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
3065..3068
/label=DUS
3076..3852
/gene="fpr"
3076..3852
/gene="fpr"
/feature="fpr"
/EC_number="1.18.1.2"
/feature="NMA1442, fpr, probable ferredoxin--NADP reductase,
len: 258 aa; similar to many e.g. SW:FERK_AZOVI
(EHML:J36319), fpr, Azolobacter vinelandii
ferredoxin--NADP reductase (EC 1.18.1.2) (257 aa), fasta
scores: E(): 0, 69.8% identity in 255 aa overlap. Similar
to NMA1664, fasta scores: E(): 1.3e-30, 34.6% identity in
257 aa overlap. Contains pfam match to entry PF00175
oxidored_fad. Oxidoreductase FAD/NAD-binding domain"
/codon_start=1
/transl_table=11
/product="putative ferredoxin--NADP reductase"
/protein_id="CAB84678.1"
/db_xref="GI:7380095"
/db_xref="SPTREMBL:Q9JRE3"
/translation="MAFNTOKVLSVHHNTDAFTTCTRDSLRPENCOPVWGLMV
DKRPLMAYVSAANNEHLEFSTIKVQDEPLSLQHLKVGVDYLISKPTGTLVAG
DLPNGKHLILSTGTIAPLSTKDEPLEDEKIIILHGVGRKKDLAYVDFTEKEL
PEHEYLGLDIKELIYPIVYSREEFHNGHGLDLMWSGKLFEDIGLPIKINPODDRAML
CGSPAMLKDTCKVLDDFGTLVSPKTVGRGDIYLERAFVQ"
3394..3768
/misc_feature

```

```

/misc_feature
/gene="fpr"
/feature="pfam match to entry PF00175 oxidored_fad,
oxidoreductase FAD/NAD-binding domain, score 20.00,
E-value 8.4e-05"
/label=DUS
3866..3875
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
3964..3983
/feature="NMA1443, len: 53 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1443"
/protein_id="CAB84679.1"
/db_xref="GI:7380096"
/db_xref="SPTREMBL:Q9JU89"
/translation="MYKKNCPMKCYAENKMKRMVFIENGCSATVIRHTGIMKRIH
IDIIIPFIA"
/complement(4509..6980)
/gene="NMA1444"
/complement(4509..6980)
/gene="NMA1444"
/feature="NMA1444, probable P-type cation-transporting
ATPase, len: 823 aa; similar to many e.g. SW:COPA_EWTHR
(EHML:J13292), copA, Enterococcus hirae
copper/potassium-transporting ATPase A (EC 3.6.1.36) (727
aa), fasta scores: E(): 0, 34.0% identity in 744 aa
overlap. Similar to NMA1539, fasta scores: E(): 0, 34.4%
identity in 735 aa overlap. Contains hydrophobic, probable
membrane-spanning regions. Contains two pfam matches to
entry PF00122 E1-E2 ATPase, E1-E2 ATPases and P500154
E1-E2 ATPases phosphorylation site"
/codon_start=1
/transl_table=11
/product="putative P-type cation-transporting ATPase"
/protein_id="CAB84680.1"
/db_xref="GI:7380097"
/db_xref="SPTREMBL:Q9JU88"
/translation="MKTKCFHGLDVPENLHLYTRENEDRETCCAGQAAVQSTIIDA
4.1%: Score 36.4; DB 1; Length 329861;
Best Local Similarity 51.2%; Pred. No. 1.4e+02;
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
Query Match
477 caaaatcatgcaagcgaggttcgagcgaagaagaaacgcgcctaccagaataca 536
11331 CGAATACGCCCTTTCAGGCTTCGGGCGCTTGGGAATATATCGCGTTACCGGTAGGAA 11390
537 aggggtcaacaataatcaaaagcctcgcttcgagcgaagaacacatcgctccgca 596
11391 TCGTGTAGACACGCGGAACAACACCATACCGCGCGCGTGTAGAGTTTGTCTCCGACCGC 11450
597 ccacggtccctccctccacgaagagcgaggaagcgatgatggtgagt 642
11451 GCCCTTCGTGTACGCGGAAAAAGCCGATGACGAAGGCGTCAT 11496
RESULT 39
AX043922 349980 bp DNA linear PAT 24-NOV-2000
LOCUS
DEFINITION Sequence 1 from Patent WO0066791.

```

ACCESSION AX043922
 VERSION AX043922.1 GI:11342850
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis.
 Neisseria meningitidis
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;

REFERENCE
 1 (bases 1 to 349980)
 AUTHORS
 Pizzo, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C., Masignani, V., Galeotti, C., Mora, M., Ratti, G., Scarcellini, M., Scarlato, V., Rappelli, R., Frazer, C.M. and Grandi, G.
 Neisseria genomic sequences and methods of their use
 Patent: WO 0066791-A 1 09-NOV-2000;
 TITLE
 JOURNAL
 CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
 FEATURES
 source
 1..349980
 Location/Qualifiers
 /organism="Neisseria meningitidis"
 /db_xref="taxon:487"
 /note="sequence too long, cut in 8 pieces. -seq 1: 1 to 349980 349980 bases-seq 108: 300001 to 649980 349980 bases-seq 109: 600001 to 949980 349980 bases-seq 110: 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to 2272325 172325 bases"

BASE COUNT 83241 a 85091 c 95206 g 86442 t
 ORIGIN

Query Match 4.1%; Score 36.4; DB 6; Length 349980;
 Best Local Similarity 44.3%; Pred. No. 1.4e+02;
 Matches 148; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

Oy 477 caaataatggaagcgaggttcgcgcaagaagaaacgcgcctccagcataca 536
 Db 115461 CAACTTTTGGCTACCTACGCGGCTACCTGCTTCACTGCGCAGCAAAACCCA 115520
 Oy 537 agggggaacaaatcatcaaacgcccgtcgtcggaagcaacatcgtccgcga 596
 Db 115521 AGGCAAGCAAAACATCAAAATCAAAAGGCGTGTGGACGCAAAAGCGCGACCG 115580
 Oy 597 ccaagtcctccctcccaagaagcgaggaagcgatgagtgatcttcggcaaac 656
 Db 115581 CCTCGCCCTGTCGCCGCCCGGCGGTCACACCATCGCAATTTCAAGGTCTGTGCA 115640
 Oy 657 tgcctatacatgacgctgagcggaacattgacacatgcaaaagcgctgaacccgtt 716
 Db 115641 GAAGCGGCAATTGAACCTGCGCGACGAATCGCGAAGTGTTCGCGTCCGAACAGAA 115700
 Oy 717 ttctgtctgcaagcctcgtcgtcggaagagtttcgattgacatccgcccgtcca 776
 Db 115701 TTGCGCCGCGCGACGCGCGGTCAAAGCGGTTTATGTATAAAAGCAACAGGCGCA 115760
 Oy 777 aggggaattgaagcggaacaaagcccatgagcc 810
 Db 115761 GACGCGGCTGAATGCCACTGCGCAAAAACC 115794

RESULT 40
 AEO01773
 LOCUS AEO01773 12514 bp DNA linear BCT 02-JUN-1999
 DEFINITION Thermotoga maritima section 85 of 136 of the complete genome.
 ACCESSION AEO01773 AEO00512
 VERSION AEO01773.1 GI:4981704
 KEYWORDS
 SOURCE Thermotoga maritima.
 ORGANISM Bacteria; Thermotogales; Thermotoga.
 REFERENCE
 1 (bases 1 to 12514)
 AUTHORS
 Nelson, K.E., Clayton, R.A., Gill, S.R., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, M.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratl, M.S., Phillips, C.A.,

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 1 (bases 1 to 12514)
 AUTHORS
 Nelson, K.E., Clayton, R.A., Gill, S.R., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, M.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratl, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
 TITLE
 JOURNAL
 Direct Submission
 Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1..12514
 Location/Qualifiers
 /organism="Thermotoga maritima"
 /db_xref="taxon:3336"
 72..461
 /gene="TM1156"
 72..461
 /gene="TM1156"
 /note="similar to GB:PYRO_h percent identity: 63.36; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AAD36237.1"
 /db_xref="GI:4981705"
 /translation="MEMKOLRSRRKNNITAGCGFAEYGVDPITVRLTWITLAW GAGLLVITLAWLMEEGESESERKPLQNTGKLVAVGLGILFGLVFLVSAFFPL FGVAMKVLVALLLVIFGLVLLRDERK"
 458..1255
 /gene="TM1157"
 458..1255
 /gene="TM1157"
 /note="similar to percent identity: 0.00; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAD36233.1"
 /db_xref="GI:4981706"
 /translation="WKILIFGLFLFGLITLTTFLDVFESISFLTRFHNGLFWPLI LITFVPIFLSVKRLMYFLSIVGSFLLILVWPYESTPGKIQVFTGKRISL KSGFTVFIIEGSEELIVASGISKISGLDIYKSGLSMRKLGKIVELPDTFEL SPENGLFTVKGFEENRFTRIETKDSVNLINPFOKMNVPYFEAEQVLEVFRLPT GSVYFIDKKDGLLKTIGNIVESDPNDFEFLKDGVRVHLEGJ"
 1257..1940
 /gene="TM1158"
 1257..1940
 /gene="TM1158"
 /note="similar to GB:AEO00782 percent identity: 63.10; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AAD36234.1"
 /db_xref="GI:4981707"
 /translation="MRVLAIRHVEIEDLGMEDIFREKNSFDYLDTPKGRLEPRL ESYLVILGNGVAGAEERKYPRIKYEPLQIEILKKEIPFGLIGSOMLAKVGSAY TRNGKEETGWTVEKVSQDNKFFRPPRLRYFQWNGDTPRLPRARVPVSEKEND GFVYKAVAGLOFHEIVGARIMKRWIEAYKDELEKKIDPRLLETAEREKVLKGLLR SLIERVMS"
 complement(2595..3539)
 /gene="TM1159"
 complement(2595..3539)
 /gene="TM1159"
 /note="similar to percent identity: 0.00; identified by sequence similarity; putative"
 /codon_start=1

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36235.1"
/db_xref="GI:4981708"
/transl_table="MEDVRLPSEYSLGYLRGKRIYSILKEDSTYSRAEKGTDPVE
VITSPKRRKQGSVMIIQGLGSONVILIMAHYLSRKGIEAILPVLPGNPTRAEES
VSGDLYSSDLDRMSRPMHALDILSLLELLKVKMKHERRNCIFGYCLGMLAVILN
ALSDFPKRTIIMAGSGDFATLPMKSPITSLFVRRLKSGKEGHEGMEENFVYVRS
LERKERSVOEMSNINPLKIDPLAAYAKFVDTSRIVMLEAMFKAALKRSTRDLIM
EHGKPKRIRVPSHVSMPQVLAARYIKLKKEOSA"
complement(3546..4466)
/gene="TM1160"
complement(3546..4466)
/note="similar to GP:2695720 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="AAD36236.1"
/db_xref="GI:4981709"
/transl_table="MYKLIFGLIITSPVLEPSNMLGLIFELLSPVYNAVGR
LPTRKSTGNTVYVYKSKSLMDVYIPSVKRESTPVLFAHGGWISGVRQPNV
SWYRELNAKFAVATDPRYRGTHTYIEDLLEDKSAISPLNENREHLIKNLMLGSL
AGHLVLYHAMRSKGEKDEGNVAVAPCDLDMWMTSLFARSVAATTLKGF
PVRRKEDYVYSPVAVNPKAPSTMLVHGKDDVYVYSSVKMYKRLRENGVAKLRL
HPGKGFEEFVLDLTVKELYETVSFLKR"
complement(4479..5819)
/gene="TM1161"
complement(4479..5819)
/note="similar to PID:1652477 percent identity: 61.31;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Mg2+ transporter MgtE, putative"
/protein_id="AAD36237.1"
/db_xref="GI:4981710"
/transl_table="MKIKVEIDLOELIEKDEFTLKVLEOODPADVKEMIEKLPPDL
KVYFRLPRDKAAEVESELEDDOLELKLPEERLKEIFESMDPPDDEVLLEMPA
NVYKLLSYLPOREBELLYILNYPEDSARALATPEYVFLFEDMYREALEKVEKAEK
KKENYISLAWIDTRKLTGVELRDILVDPQKYSLEINKDPVYHAIDDELAEK
MKKYDILITPVYDSEERLVTTFEDLVVIEEATEDIOKMASNATYSYFHPNAP
KILKRSPIVVLLESYNGNITSSYEFPLASIPILAFITMISAGNTAQVLSAL
MIRGFPLNEISLKMVKVLLRESLIGSTGLIAGVLYRAFLISSDPLINFAQTAL
LVILYANIMGMLPFIARIFKIDPAFMAGPLITIVDTGIMITYVYVHSPLS"
6027..6707
/gene="TM1162"
6027..6707
/note="similar to GB:AE000782 percent identity: 67.69;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36238.1"
/db_xref="GI:4981711"
/transl_table="MKVTFGLHAVVLIIEGKNIIIDPISGNPCVPKLEGLPKIDYI
LVTHGHDHLAGAVEIAKKNDATVINSNEICHLGKRGVKTAMHIGSYLDFGRVK
MTPAVHSGIILDGSMYIGGNPSGLITIEGKRIYHAGGTGLTRMELIAEENNVAF
LPIGNFVMDVEDAVRAAVMIKPKVPMHYGTWMLIRADVELFKKKKEKGVCEVIL
EPGESLEL"
6720..8192
/gene="TM1163"
6720..8192
/note="similar to GB:AE000657 percent identity: 57.24;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, GGDEF domain"
/protein_id="AAD36239.1"
/db_xref="GI:4981712"

```

```

/transl_table="MOVFFYVVSFGELVAFALYLFCHVKKRIEYRRLKESDITDGL
SKIDPSPPEEFFQKILNLAIVEVPEAVGVSRTIPSGDMJFVRSRHIYDLGKRF
PSRYLFRAGRDVEVNFIDYDRLDPDDMKFEETLGTGKSLVYGJFSGDEFVNTI
ALDSPYSGFSELSKKVLSKLSGLASTYLLMRKTLNBOOLKVMSTITLLLRKSI
SVEEFLEAKLQIVESANVETGIVPEGRIPEPHDLEIPEPDESEKBEKIEKLY
KGVHYTHVLDIKDMPSSHYMFLSNEDILARSFVSUWTFREVSUJLREYHLQKTR
ELAKDPLTEASRRTFNEMFTSHAMLRNOKRSVLIIMVDGLKMLRNDYHGLMGD
KALVEVTKLKRTRFRESDLVFRYGGDEELVLVYSTKNMAASVYERVENLKDVLDP
KIEFSGYEEDIGFMPRIERALRADLLYKNNFEKRGSGSE"
8189..9865
/gene="TM1164"
8189..9865
/note="similar to GB:AE000666 percent identity: 68.99;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="2-oxoacid ferredoxin oxidoreductase, alpha
subunit"
/protein_id="AAD36240.1"
/db_xref="GI:4981713"
/transl_table="MKDVSIVLSCGAGGIGIOTVENVLRVLKDSGFHFATREKMSRY
RGNNWTEIRVSSRRVRSFVDRIDVLPLGAGATRLKRLRITKTIYVIGEEFTIKESP
GEKIVIPFLNTAKOIKSIVANSVAILGSLGALNEDALDOITRQSSGEDIYND
NIRALIEYKKGKGOELSOKEIPEVDEDSIKNEVLINGEAYAGLGLIAGCNFVSYPM
SPSTVILVFLAOKHDKDFIVVQADEDEIAANMILGAYAGRLVYTSGGFVSSYPM
ALSLAGMIESPAVNHIAQRPGPATGLPRTDQGLNLALVAGHGDFFPVIYAPQNVIE
AFYLQKAEVNAADKYQVPEFVLTQYLVDSYNNLPGLDELKYEKHIKTRDYIRY
AITEDGISPRGIPGYEGELVRYVDSDEHDEFCHITDEPNTRYAVNKRRLRGKETLKEI
VAPKLIQDENRVLVWAGSTLEPIKEALIEGLDVALILHFSQWPIDESVAITYEKKAE
KVVAVEGNATGOFANLIRQVYGFHKDKILKYNKIQPSVLELKKIAEVL"
9866..10714
/gene="TM1165"
9866..10714
/note="similar to GB:AE000666 percent identity: 77.82;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="2-oxoacid ferredoxin oxidoreductase, beta
subunit"

```

Query Match 4.0%; Score 36.2; DB 1; Length 12514;
Best Local Similarity 45.8%; Pred. No. 1.3e+02;
Matches 125; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

```

QY 255 ggaacttgcgccgcytttttcagaagaacccggaagacacacacatgtcaaaagcgt 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1841 GAAGATAGATCCAGGCTTCCTTGAAACGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1900

QY 315 aacggcgttggaacatgtgacgacgctttgacacaaacacgaaggcgtcatcatcac 374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1901 CCTCTTGAAGGCTCTTTTGGAAAGATGAGTGGAAAGCTGAGAGGCTGTTATTGGATACC 1960

QY 375 gccgacacatcggaacatagatttggcggaagcgtacacacgacgacgcttcgtccc 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1961 TCTCATCATCCACATCCATCCGATTTCAACAGGAATTACAGAAACACGCGATCTTAATATC 2020

QY 435 gctgacccgcacatgatacaaacccggaacacacacacacacacacacacacacacac 494
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2021 TCTGCTGAGTTTGTGTTTCCAAAAAATGAGAGCATGAGTACATGATGCTTCCCGA 2080

QY 495 cagggttcgcgcaagaagaaacacgacacacacacacacacacacacacacacacacacac 527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2081 AAGATTCCTGAAATTGAAACATCTCTTCCTTC 2113

RESULT 41
AC018567 176106 bp DNA linear HTG 26-MAY-2000
LOCUS Homo sapiens clone Rpl1-115H16, WORKING DRAFT SEQUENCE, 24
DEFINITION uncloned pieces.
ACCESSION AC018567

```


VERSION AC018567.4 GI:8072480
 KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 176106)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens, clone RP11-115H16
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 176106)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, R.,
 Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearlano, K., Dewar, K., Domino, M., Doyle, M., Fennestor, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Garayna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehocsky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 Melidrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H.,
 O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A.,
 Santos, R., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,
 Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A.
 and Zody, M.
 Direct Submission
 Submitted (14-DEC-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced g1:7341730.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WtBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1873
 Center clone name: 115_H.16
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Sequencing vector: Plasmid; n/a; %0.1% of reads
 0.257234726688103chemistry: Dye-terminator Big Dye; 100% of
 reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 160129 bases at least Q40
 Consensus quality: 168013 bases at least Q30
 Consensus quality: 171246 bases at least Q20
 Insert size: 158000; agarose-ff
 Insert size: 173806; sum-of-coverage
 Quality coverage: 4.2 in Q20 bases; agarose-ff
 Quality.
 NOTE: This is a 'working draft' sequence. It currently
 consists of 24 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 * 1 1492: contig of 1492 bp in length
 * 1493 1592: gap of 100 bp
 * 1593 3341: contig of 1749 bp in length
 * 3342 3441: gap of 100 bp
 * 3442 5215: contig of 1774 bp in length
 * 5216 5315: gap of 100 bp
 * 5316 6282: contig of 967 bp in length
 * 6283 6382: gap of 100 bp
 * 6383 8157: contig of 1775 bp in length

FEATURES
 source
 * 8158 8257: gap of 100 bp
 * 8258 11448: contig of 3191 bp in length
 * 11449 11548: gap of 100 bp
 * 11549 14787: contig of 3239 bp in length
 * 14788 14887: gap of 100 bp
 * 14888 18631: contig of 3744 bp in length
 * 18632 18731: gap of 100 bp
 * 18732 23654: contig of 4923 bp in length
 * 23655 23754: gap of 100 bp
 * 23755 27602: contig of 3848 bp in length
 * 27603 27702: gap of 100 bp
 * 27703 32499: contig of 4797 bp in length
 * 32500 32599: gap of 100 bp
 * 32600 38551: contig of 5952 bp in length
 * 38552 38651: gap of 100 bp
 * 38652 46281: contig of 7630 bp in length
 * 46282 46381: gap of 100 bp
 * 46382 54455: contig of 8074 bp in length
 * 54456 54555: gap of 100 bp
 * 54556 62059: contig of 7504 bp in length
 * 62060 62159: gap of 100 bp
 * 62160 71289: contig of 9130 bp in length
 * 71290 71389: gap of 100 bp
 * 71390 78460: contig of 7071 bp in length
 * 78461 78560: gap of 100 bp
 * 78561 89295: contig of 10735 bp in length
 * 89296 89395: gap of 100 bp
 * 89396 101950: contig of 12555 bp in length
 * 101951 102050: gap of 100 bp
 * 102051 113969: contig of 11919 bp in length
 * 113970 114069: gap of 100 bp
 * 114070 128784: contig of 14715 bp in length
 * 128785 128884: gap of 100 bp
 * 128885 144042: contig of 15158 bp in length
 * 144043 144142: gap of 100 bp
 * 144143 157099: contig of 12957 bp in length
 * 157100 157199: gap of 100 bp
 * 157200 176106: contig of 18907 bp in length.
 Location/Qualifiers
 1. 176106
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-115H16"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1492
 /note="assembly-fragment"
 /note="3341
 /note="assembly-fragment"
 3442. 5215
 /note="assembly-fragment"
 5316. 6282
 /note="assembly-fragment
 clone_end:17
 vector_side:left"
 6383. 8157
 /note="assembly-fragment"
 8258. 11448
 /note="assembly-fragment"
 11549. 14787
 /note="assembly-fragment"
 14888. 18631
 /note="assembly-fragment"
 18732. 23654
 /note="assembly-fragment"
 23755. 27602
 /note="assembly-fragment"
 27703. 32499
 /note="assembly-fragment"
 32600. 38551
 /note="assembly-fragment"
 38652. 46281
 /note="assembly-fragment"
 46382. 54455

	Query Match	4.0%	Score 36.2	DB 2	Length 176106
	Best Local Similarity	59.0%	Pred. No. 1.5e+02		
	Matches 62	Conservative	0	Mismatches 43	Indels 0
				Gaps	0
QY	520	gagcctaccagcagctaaagggtgtaacaatactcaagcccgcttcggcggaagca	579		
Db	69398	GCTTCCAGATTACACAAAGGGTCTCAAAATTCAACCTGTTCTTAAAGTTGGGCCATTCC	69457		
QY	580	acatcagtcctcgccggaaccagctccctccctccctaaaggcgcg	624		
Db	69458	ACCCCTCTTAATGTCCTTCCCAACCCCTGACCTGTATGACAAAGGG	69502		

```

TITLE
JOURNALML
Direct Submission
Submitted (24-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 1, 2002 this sequence version replaced gi:18378741. All repeats were identified using RepeatMasker:
Smtt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

--- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 12637
Center clone name: 418_K_9

NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submittor.
This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
1 176338: contig of 176338 bp in length
1 176339 176438: gap of 100 bp
176439 185139: contig of 8701 bp in length.

FEATURES
Source
Location/Qualifiers
1..185139
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18"
/clone_lib="RP11-418K9"
/cloned_lib="RCRT-11 Human Male BAC"

BASE COUNT      47226 a 44480 c 43992 g 49341 t    100 others
ORIGIN
Query Match          4.0%   Score 36.2; DB 2; Length 165139;
Best Local Similarity 59.0%; Pred. NO. 1.5e+02;
Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 520 gcgcctaccagcacaagaagggtcaaaacatcatcaaagccgcttcggcggaagca 579
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 168564 GCATCCAGATAAACAAGAAGGGGTCTCAGATTGCAAGCGTCTGAATTGGGCGCATTC 168623
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 580 accattgctctgccgcgacgaagtccccctcccctaagaagcgagg 624
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 168624 ACCCTCTTTATGTCCTTCCCACCCTGCCTTGATGCCAACAAAGG 168668
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 43
AC021934
LOCUS AC021934 185823 bp DNA linear HTG 16-MAR-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-313C14 map 18, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC021934
VERSION AC021934.3 GI:7249154
KEYWORDS HTG: HTGS_PHASEI; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185823)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-313C14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185823)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barra,N., Beckerly,R., Bedalov,F.,

```

TITLE
JOURNAL
COMMENT

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choele, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArliano, K., Dewar, K., Domino, M., Doyle, M., Fenebor, J.,
Fertile, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, J., McEwan, P., McGurk, A., McKernan, K.,
Mepheters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,
Piette, N., Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A., and Zody, W.

Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced g1:6970535.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L5212
Center clone name: 313_C-14

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 178071 bases at least Q40
Consensus quality: 161967 bases at least Q30
Consensus quality: 183256 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 184223; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 17 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1742: contig of 1742 bp in length
* 1743 1842: gap of 100 bp
* 1843 5170: contig of 3328 bp in length
* 5171 5270: gap of 100 bp
* 5271 7808: contig of 2538 bp in length
* 7809 7908: gap of 100 bp
* 7909 11875: contig of 3967 bp in length
* 11876 11975: gap of 100 bp
* 11976 15932: contig of 3957 bp in length
* 15933 16032: gap of 100 bp
* 16033 21963: contig of 5931 bp in length
* 21964 22063: gap of 100 bp
* 22064 29014: contig of 6951 bp in length
* 29015 29114: gap of 100 bp
* 29115 35138: contig of 6024 bp in length
* 35139 35238: gap of 100 bp
* 35239 42525: contig of 7287 bp in length
* 42526 42625: gap of 100 bp
* 42626 51181: contig of 8556 bp in length
* 51182 51281: gap of 100 bp
* 51282 63207: contig of 11926 bp in length
* 63208 63307: gap of 100 bp
* 63308 76220: contig of 12913 bp in length

FEATURES
source
* 76221 76320: gap of 100 bp
* 76321 87341: contig of 11021 bp in length
* 87342 87441: gap of 100 bp
* 87442 103490: contig of 16049 bp in length
* 103491 103590: gap of 100 bp
* 103591 121791: contig of 18201 bp in length
* 121792 121891: gap of 100 bp
* 121892 149125: contig of 27234 bp in length
* 149126 149225: gap of 100 bp
* 149226 185823: contig of 36598 bp in length.
Location/Qualifiers
1. 185823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-313C14"
/clone_11b="RCR1-11 Human Male BAC"
1. 1742
/note="assembly_fragment"
1843. 5170
/note="assembly_fragment"
5271. 7808
/note="assembly_fragment"
7909. 11875
/note="assembly_fragment"
11976. 15932
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
16033. 21963
/note="assembly_fragment"
22064. 29014
/note="assembly_fragment"
29115. 35138
/note="assembly_fragment"
35239. 42525
/note="assembly_fragment"
42626. 51181
/note="assembly_fragment"
51282. 63207
/note="assembly_fragment"
63308. 76220
/note="assembly_fragment"
76321. 87341
/note="assembly_fragment"
87442. 103490
/note="assembly_fragment"
clone_end:r7
vector_side:right"
103591. 121791
/note="assembly_fragment"
121892. 149125
/note="assembly_fragment"
149226. 185823
/note="assembly_fragment"

BASE COUNT 49638 a 43397 c 41719 g 49467 t 1602 others
ORIGIN

Query Match 4.0%; Score 36.2; DB 2; Length 185823;
Best Local Similarity 59.0%; Pred. No. 1.5e+02;
Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 520 ggcgcctaccagcgcatacaaggggtcaacaacatcatcaagccctgcgttcgaggaagca 579
DB 46296 GCTCCAGCAATACACAAAGGCTCAGCAATTCAGCTCTCTAAGTTGGGGCCATTCC 46355
OY 580 acatcgtctcgcgcacacagctccctccctcccaagaagcgagg 624
DB 46356 ACCCTTTATGTCCTTCCACCCCTGCTGATGCACAAAGG 46400

RESULT 44
AC026740 194763 bp DNA linear HTG 31-AUG-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2589H19, WORKING DRAFT
DEFINITION AC026740.5 GI:15290380
ACCESSION AC026740.5
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 194763)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194763)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:13699695.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 824344
Center clone name: CTD-EL_2589H19

Summary Statistics
Consensus quality: 189571 bases at least Q40
Consensus quality: 192971 bases at least Q30
Consensus quality: 194072 bases at least Q20
Estimated insert size: 221540; agarose-1p estimation
Quality coverage: 6.01 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.84 in Q20 bases; agarose-1p estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 2959: contig of 2959 bp in length
* 2960 3059: gap of unknown length
* 3060 7889: contig of 4830 bp in length
* 7890 7989: gap of unknown length
* 7990 30637: contig of 22648 bp in length
* 30638 30737: gap of unknown length
* 30738 194763: contig of 164026 bp in length.
Location/Qualifiers
1. 194763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2589H19"
/clone_1lb="Caltech human BAC library D"
BASE COUNT 45553 a 50681 c 51512 g 46717 t 300 others
ORIGIN
Query Match 4.0%; Score 36.2; DB 2; Length 194763;
Best Local Similarity 50.3%; Pred. No. 1.5e+02;
Matches 89; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 415 agcagcagcttcgcgtccgcgtgacgcgcgtacacacgcgcgaataacgaagcgtata 474
Db 17070 AGCAAGCTGCTTACACCTGAGTACGCTTACGCTTCTCTCACAACTGCTTCA 17129

QY 475 gacaaatcatgcagggcgaggttcgcgcgaaggaacacgcgcctaccagcata 534
Db 17130 GAGAGGCACCTCCCAAAAGCTGACAGCAAGAAAGGATATGAGGCATCTATCTATCTCG 17189
QY 535 caaggggtcaacaaatcatcgaagccgtgcgttcggcggaagcaccatgcgtctc 591
Db 17190 AAAGCATGATTCATACAAAGAAAGCCCTATGATCGGCAAAACATCAAAAGCCCTG 17246
RESULT 45
AC026740/c 194763 bp DNA linear HTG 31-AUG-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2589H19, WORKING DRAFT
DEFINITION AC026740.5 GI:15290380
ACCESSION AC026740.5
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 194763)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194763)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:13699695.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 824344
Center clone name: CTD-EL_2589H19

Summary Statistics
Consensus quality: 189571 bases at least Q40
Consensus quality: 192971 bases at least Q30
Consensus quality: 194072 bases at least Q20
Estimated insert size: 221540; agarose-1p estimation
Quality coverage: 6.01 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.84 in Q20 bases; agarose-1p estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 2959: contig of 2959 bp in length
* 2960 3059: gap of unknown length
* 3060 7889: contig of 4830 bp in length
* 7890 7989: gap of unknown length
* 7990 30637: contig of 22648 bp in length
* 30638 30737: gap of unknown length
* 30738 194763: contig of 164026 bp in length.
Location/Qualifiers
1. 194763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2589H19"
/clone_1lb="Caltech human BAC library D"
BASE COUNT 45553 a 50681 c 51512 g 46717 t 300 others
ORIGIN
FEATURES
source

